SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(134th Edition)

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

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

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

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

Central Library

CONTENTS

1. A 3D-convolutional-autoencoder embedded Siamese-attention-network for classification of hyperspectral Images, ***6.Pallavi Ranjan***, ***3.Rajeev Kumar*** and Ashish Girdhar, CSE, DTU

2. A Deep Learning Model to Generate Image Captions, ***8.Shashank Parmar***, ***8.Raman Tyagi*** and ***8.Prince Kumar Dhankar***, Software Engineering, DTU

3. A human activity recognition framework in videos using segmented human subject focus, ***6.Shaurya Gupta***, ***3.Dinesh Kumar Vishwakarma*** and ***3.Nitin Kumar Puri***, Applied Physics and IT, DTU

4. A Systematic and Comprehensive Review on 2-D and 3-D Numerical Modelling of Stirling Engine, ***8.Vaibhav Singh*** & ***3.Anil Kumar***, Mechanical & Environment, DTU

5. A Systematic Literature Review on Maintenance of Software Containers, ***3.RUCHIKA MALHOTRA***, ***6.ANJALI BANSAL*** and MAROUANE KESSENTINI, CSE, DTU

6. A systematic literature survey on skin disease detection and classifcation using machine learning and deep learning, ***6.Rashmi Yadav*** and ***3.Aruna Bhat***, CSE, DTU

7. Advanced Intelligent Hybrid Approach for State of Charge Estimation of Li-Ion Batteries in Electric Vehicles under Various Drive Cycles, ***6.Deepak Kumar***, ***3.M. Rizwan*** and ***3.Amrish K. Panwar***, Electrical and Applied Physics, DTU

8. Advancements in composite cathodes for intermediate-temperature solid oxide fuel cells: A comprehensive review, Anil Kumar Yadav, Shailendra Sinha and ***3.Anil Kumar***, Mechanical and Environment, DTU

9. Altered DTI scalars in the hippocampus are associated with morphological and structural changes after traumatic brain injury, Palkin Arora, Richa Trivedi, ***6.Megha Kumari***, Kavita Singh, Rajat Sandhir, Maria M. D'Souza and Poonam Rana, Biotechnology, DTU

10. An Examination of A Single-Pass Heat Exchanger's Parametric Design Using Coding, Pardeep kumar Monu, ***3.Naveen Kumar***, Raj Kumar Raj, Amit Sharma Amit and Anju Anju, Mechanical, DTU

11. An Integrated Entropy-TOPSIS Approach for Electric Vehicle Selection, Kiran Pal, Dhruv Saraswat and ***6.Neeraj Budhraja***, Mechanical, DTU

12. Analyzing Barriers in Adoption of Artificial Intelligence for Resilient Health Care Services to Society, ***3.Girish Kumar***, Rajesh Kumar Singh, Vedpal Arya and ***8.Shivam Kumar Mishra***, Mechanical, DTU

13. Analyzing the Performance of Class-F Power Amplifier by Comparing Microstrip and SSL/SIL Transmission Lines Based Matching Networks for Smart Parking Solutions for Improved Efficiency and Output Power, ***6.Puja Priya*** and ***3.Gurjit Kaur***, Electronics, DTU

14. Approximation in scour depth around spur dikes using novel hybrid ensemble data-driven model, ***6.Balraj Singh*** and ***3.Vijay K. Minocha***, Civil, DTU

15. Automatic Construction of Interval-Valued Fuzzy Hindi WordNet using Lexico-Syntactic Patterns and Word Embeddings, ***3.MINNI JAIN***, ***3.RAJNI JINDAL*** and AMITA JAIN, CSE, DTU

16. Bond Graph Simulation of Composite Rotor's Nonlinear Response, ***8.Anuj Sharma***, ***3.Atul Kumar Agrawal***, Shadab Ahmad, Shanay Rab, and Yebing Tian, Mechanical and Industrial and Automobile, DTU

17. CeO2 and PEDOT:PSS modifed conducting paper for organophosphate pesticide detection, ***6.Saroj Paneru***, ***6.Sweety*** and ***3.Devendra Kumar***, Applied Chemistry, DTU

18. Chopped carbon fibre reinforcement in high-strength concrete: Flexural characteristics, ***6.Nitin Lamba***, ***3.Ritu Raj*** and ***3.Poonam Singh***, Civil and Applied Chemistry, DTU

19. Coal power plant-enabled grid resilience through distributed energy resources and demand response integration, ***6.Vivek Saxena***, ***3.Narendra Kumar*** and ***3.Uma Nangia***, Electrical, DTU

20. Communication independent Islanding Detection technique utilizing Wavelet Packet Transform, ***7.Samavedam Swetha*** and ***3.Vishal Verma***, Electrical, DTU

21. Deployment and performance monitoring of docker based federated learning framework for software defect prediction, ***3.Ruchika Malhotra***, ***6.Anjali Bansal*** and Marouane Kessentini, Software Engineering, DTU

22. Development of KNNLTS–PVDF-based flexible piezoelectric generator for energy-harvesting application, ***7.KOMAL VERMA*** and ***3.RICHA SHARMA***, Applied Physics, DTU

23. Device Optimization of T-shaped Gate and Polarized Doped Buffer-Engineered InAlN/GaN HEMT for Improved RF/Microwave Performance, ***6.Megha Sharma*** and ***3.Rishu Chaujar***, Applied Physics, DTU

24. Effect of Multi-step Sintering on the density of Li7La3Zr1.75Ce0.25O12 as Solid State Electrolyte Material, ***6.Sharad Singh Jadaun***, ***3.Amrish K Panwar*** and Geetanjali, Applied Physics, DTU

25. Effect of Rare Earth Metal (Lanthanum) Doping on Optical and Electronic Properties of Lithium Niobate (LiNbO3): DFT Insights, ***6.Ashish Raturi***, ***3.Poornima Mittal*** and Sudhanshu Choudhary, Electronics, DTU

26. Efficient prediction of coronary artery disease using machine learning algorithms with feature selection techniques, Md. Mehedi Hassan, Sadika Zaman, Md. Mushfiqur Rahman, Anupam Kumar Bairagi, Walid El-Shafai, Rajkumar Singh Rathore and ***7.1.Deepak Gupta***, CSE, DTU

27. Emission performance evaluation of urban passenger car feet through feld investigation data in a megacity, Abhinav Pandey, Govind Pandey and ***3.Rajeev Kumar Mishra***, Environmental, DTU

28. Energy transfer mechanisms and color-tunable luminescence of Tm3+/Tb3+/Eu3+ co-doped Sr4Nb2O9 phosphors for high-quality white lightemitting diodes, Ravina, Kanishk Poria, Mukesh K. Sahu, A. Kumar, ***6.Anu***, Sajjan Dahiya, Nisha Deopa and ***3.A. S. Rao***, Applied Physics, DTU

29. Energy-Efcient Fuzzy Scheduling System for Crankcase Covers Manufacturing, Sumit Chawla & ***3.Ranganath M. Singari***, Mechanical, Production, & Industrial, DTU

30. Enhancing Quality Management in Ethiopia's Higher Education through Industry 4.0 Technologies, ***6.Asalf Habtegeorgis***, ***3.G.C Maheshwari*** and ***3.Rajan Yadav***, DSM, DTU

31. Facile synthesized Sb2S3 based high-performance visible photodetector for weak optical signal detection, Pukhraj Prajapat, Pargam Vashishtha, Deependra Kumar Singh, Rajat Kumar Mukherjee, Jai Tawale, ***6.1.Lalit Goswami*** and Govind Gupta, Electronics, DTU

32. Gravitational Lensing and Clues of rd and H0 Tension: Viscous Modified Chaplygin Gas and Variable Modified Chaplygin Gas in Loop Quantum Cosmology, Rownak Kundu, Ujjal Debnath, ***6.Himanshu Chaudhary*** and G. Mustafa, Applied Mathematics, DTU

33. Greenfield settlements due to tunnelling using tunnel boring machine (TBM) in layered soils: a parametric study, ***6.JYOTI AGARWAL*** and ***3.RAJU SARKAR***, Civil, DTU

34. Harnessing AI and Gut Microbiome Research for Precision Health, Ritcha Saxena, ***8.Vikas Sharma***, Ananya Raj Saxena and Aakash Patel, CSE, DTU

35. HOW INDIA ACHIEVES SPACE MISSION SUCCESS ON A SHOESTRING BUDGET: A LESSON IN ECONOMIC EFFICIENCY, ***7.Lubhani Kapoor*** and Dr. Sayak Bhattacharya, Economic, DTU

36. Impact of external magnetic field on temperature anisotropy driven Weibel instability in a magnetized plasma, Twinkle Pahuja, Amit Kumar, ***6.Himank Sagar***, Ruby Gupta and Jyotsna Sharma, Applied Physics, DTU

37. Investigation and mitigation of unbalanced hall sensor signal faults in sensored brushless DC motor drives, ***6.Arvind Goswami***, ***3.Mini Sreejeth***, ***3.Madhusudan Singh***, Electrical, DTU

38. Investigation of a Gate Stack Gate-All-Around Junctionless Nanowire Field-Efect Transistor for Oxygen Gas Sensing, ***3.Rishu Chaujar*** and Mekonnen Getnet Yirak, Applied Physics, DTU

39. Investigation of wire EDM process parameter on kerf width for high chromium high carbon steel by response surface methodology, Daniel Gnanaprakasam, ***3.Girish kumar*** and Suhail Ahmad Siddiqui, Mechanical, DTU

40. IoT-driven augmented reality and virtual reality systems in neurological sciences, ***6.Mehar Sahu***, ***6.Rohan Gupta***, ***6.Rashmi K. Ambasta*** and ***3.Pravir Kumar***, Biotechnology, DTU

41. Localization of beam generated whistler wave and turbulence generation in reconnection region of magnetopause, ***6.Jyoti***, ***3.Suresh. C. Sharma*** and R. P. Sharma, Applied Physics, DTU

42. Low-Energy Decentralized Context-aware Access Control in Internet of Things, ***7.1.Surendra Tyagi***, Devesh Jinwala and Subhasis Bhattacharjee, CSE, DTU

43. Measurement of Mechanical Properties of Friction Stir Welding of Al– Li alloy under different environmental conditions, ***6.S. Singh***, ***3.N. Yuvaraj*** and ***3.R. Wattal***, Mechanical, DTU

44. Microplastics influence the functional responses of a tropical estuarine calanoid Pseudodiaptomus annandalei, ***3.Jaigopal Sharma***, Malayaj Rai and Robert S. Guino-o II, Biotechnology, DTU

45. Modeling and simulation of FIR flter using distributed arithmetic algorithm on FPGA, Amrita Rai, Ajay Roy, Shamimul Qamar, Abdulelah G. F. Saif, Magdi Mohammad Hamoda, ***6.Abdul Azeem*** and Salman Arafath Mohammed, Electrical, DTU

46. Novel hybrid-CMOS inverter utilizing phase transition material for enhancing digital logic performance at lower operating voltages, ***6.Ritika Sorot***, Anubha Goel and ***3.Sonam Rewari***, Electronics, DTU

47. Optimization of acidic pre-treatment conditions using response surface methodology for ethanol production from Pistia stratiotes using Saccharomyces cerevisiae and Pichia stipites, ***6.Sweeti Mann***, ***3.Jai Gopal Sharma*** and ***3.1.Rashmi Kataria***, Biotechnology, DTU

48. Performance Parameter Extraction of SnZnTe Absorber Layer Photovoltaic Cell, ***6.Paritosh Chamola*** and ***3.Poornima Mittal***, Electronics, DTU

49. Phytochemical, antioxidant and antimicrobial activity of Livistona chinensis (Jacq.) R.Br. ex Mart. Seeds, ***7.Tamanna Beniwal***, ***6.Ritu Sharma*** and ***3.Rajinder K Gupta***, Applied Chemistry, DTU

50. Protective effects of fecal microbiota transplantation against ischemic stroke and other neurological disorders: an update, Tousif Ahmed Hediyal, C. Vichitra, Nikhilesh Anand, Mahendran Bhaskaran, Saeefh M. Essa, ***3.Pravir Kumar***, M. Walid Qoronfleh, Mohammed Akbar, Ruchika Kaul-Ghanekar, Arehally M. Mahalakshmi, Jian Yang, Byoung-Joon Song, Tanya M. Monaghan, Meena Kishore Sakharkar and Saravana Babu Chidambaram, Biotechnology, DTU

51. Quantum Image processing algorithms comparison using datasets of machine learning based applications, ***6.BARKHA SINGH***, ***3.SreeDevi Indu*** and ***3.Sudipta majumdar***, DTU

52. Retraction Note: Modelling product lines diffusion: a framework incorporating competitive brands for sustainable innovations, ***3.Deepti Aggrawal***, Adarsh Anand, Gunjan Bansal, Gareth H. Davies, Parisa Maroufkhani and Yogesh K. Dwivedi, USME, DTU

53. Role of nonlinear structures and associated turbulence generation dayside magnetosphere reconnection sites, ***6.Neha Pathak***, ***3.Suresh C. Sharma***, Garima Patel and R. P. Sharma, Applied Physics, DTU

54. Seagull optimization algorithm for node localization in wireless sensor networks, ***6.Yogendra Mohan***, ***3.Rajesh Kumar Yadav*** and Manisha Manjul, CSE, DTU

55. Securing cloud-enabled smart cities by detecting intrusion using sparkbased stacking ensemble of machine learning Algorithm, ***6.Mohd. Rehan Ghazi*** and ***3.N. S. Raghava***, Electronics, DTU

56. Soil-slope stability investigation using different nail inclinations: a comprehensive LSD, FEM and experimental approach, ***6.PRASHANT C RAMTEKE*** and ***3.ANIL KUMAR SAHU***, Civil, DTU

57. Tribological Behaviour of Austenitic Stainless Steel-Clad Surface Over Low Carbon Steel Produced by Cold Metal Transfer Welding Process, ***6.Varsha Mishra***, ***3.N. Yuvaraj*** and ***3.Vipin***, Mechanical, DTU

58. Validation of a Parameter Design methodology for Air-core two-coil WPT system with multi-resonant LC compensation, ***8.Akshay***, ***8.Gaurav Yadav*** and ***3.Mukhtiar Singh***, Electrical, DTU

59. VGGIN-NET: Deep Transfer Network for Imbalanced Breast Cancer Dataset, ***6.Manisha Saini*** and ***3.Seba Susan***, CSE and IT, 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*

Neural Computing and Applications
https://doi.org/10.1007/s00521-024-09527-y

ORIGINAL ARTICLE

A 3D-convolutional-autoencoder embedded Siamese-attention-network for classification of hyperspectral images

Pallavi Ranjan[1] · Rajeev Kumar[1] · Ashish Girdhar[2]

Received: 11 July 2023 / Accepted: 15 January 2024
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2024

Abstract
The classification of hyperspectral images (HSI) into categories that correlate to various land cover sorts such as water bodies, agriculture and urban areas, has gained significant attention in research due to its wide range of applications in fields, such as remote sensing, computer vision, and more. Supervised deep learning networks have demonstrated exceptional performance in HSI classification, capitalizing on their capacity for end-to-end optimization and leveraging their strong potential for nonlinear modeling. However, labelling HSIs, on the other hand, necessitates extensive domain knowledge and is a time-consuming and labour-intensive exercise. To address this issue, the proposed work introduces a novel semi-supervised network constructed with an autoencoder, Siamese action, and attention layers that achieves excellent classification accuracy with labelled limited samples. The proposed convolutional autoencoder is trained using the mass amount of unlabelled data to learn the refinement representation referred to as 3D-CAE. The added Siamese network improves the feature separability between different categories and attention layers improve classification by focusing on discriminative information and neglecting the unimportant bands. The efficacy of the proposed model's performance was assessed by training and testing on both same-domain as well as cross-domain data and found to achieve 91.3 and 93.6 for Indian Pines and Salinas, respectively.

Keywords Classification · Deep learning · Siamese · Hyperspectral images · Autoencoder · Attention

1 Introduction

Hyperspectral imaging is an advanced method for capturing highly detailed images of the Earth's surface, allowing for high spectral resolution [1]. These images offer valuable insights into the composition of objects in the scene. The classification of these objects, which entails stratifying the pixels in the hyperspectral image (HSI), is an essential process in various fields, such as remote sensing, computer vision, environmental monitoring, and resource management [2]. Various deep learning models have been utilized for HSI classification [3], including both supervised and semi-supervised, such as autoencoders (AE) [4], convolutional neural networks (CNN) [5], long short-term memory (LSTM) [6], transfer learning (TL) [7], deep belief network (DBN) [8], have been introduced as these models have data-driven feature learning capabilities. Among the aforementioned classes of models, the supervised networks [9–11] have evolved to achieve almost perfect classification accuracy. Nevertheless, the process of obtaining labelled samples continues to pose a challenge in HSI classification [12], as the labelling remains a manual, time-consuming, and labour-intensive task, which necessitates the expertise of individuals to identify and label each pixel in the image.

In contrast, semi-supervised models require less manual data labelling, which saves time and money [13, 14], however, their performance has not been at par with the supervised models. To address this, researchers like [15]

✉ Rajeev Kumar
rajeevkumar@dtu.ac.in

Pallavi Ranjan
pallaviranjan_2k20phdco511@dtu.ac.in

Ashish Girdhar
ashishgirdhar@kuk.ac.in

[1] Department of Computer Science, Delhi Technological University, Delhi, India

[2] Department of Computer Applications, Kurukshetra University, Kurukshetra, India

Published online: 20 February 2024

incorporate various deep learning methods, particularly few-shot learning (FSL) into hyperspectral semi-supervised models to make notable advancements in dealing with the limited data challenge and also to improve the classification accuracy. Models like these enable the utilization of a few labelled samples to learn generalized representations and improve classification accuracy on unseen data points [16]. In other words, the integration of FSL models like Siamese into hyperspectral semi-supervised models provides a promising avenue to address the limited data challenge and enhance classification performance in hyperspectral data analysis [17]. Building upon these advancements, this paper aims to further improve the performance of semi-supervised models by integrating the strengths of few-shot learning and Siamese networks not only on same-domain but also on cross-domain datasets. More specifically, the objective is to tackle the challenges posed by limited data and enhance the classification accuracy, while also taking into account concerns, such as overfitting and gradient vanishing. The contribution of the proposed work can be summarized in the following points

- The proposed network distinctively amalgamates 3D convolutional autoencoder, Siamese network, and attention layers-setting it apart for superior performance in cross-domain and same-domain hyperspectral image (HSI) classification, substantiating its novelty through experimental contributions. The crux of our innovation is the seamless integration and synergy of key components-3D convolutional autoencoder, Siamese network, and attention layers-setting our approach apart for enhanced performance in both cross-domain and same-domain HSI classification.
- Our model's semi-supervised nature is a pivotal strength, effectively overcoming the challenge of limited labelled data in HSI. This aspect is particularly valuable for real-world applications where obtaining labelled data is often resource-intensive. The 3D-CAE significantly contributes to unsupervised feature learning, generating features that may lack class separability but play a crucial role in the overall effectiveness of the model.
- The Siamese network backed up with binary cross entropy adds value by enhancing feature separability, leveraging three-dimensional convolutions to reduce intraclass distances and increase interclass distances.
- To efficiently learn with the available HSI three-dimensional hyperspectral data, the embedded attention layers in the network, which include both spectral and spatial, reduce the duplicate information by emphasizing only significant channels and pixels and suppressing the less informative ones.
- The effectiveness of the proposed semi-supervised spectral–spatial network is verified through experiments conducted on benchmark hyperspectral image (HSI) datasets. The results demonstrate that the network achieves exceptional overall accuracy, not only on same-domain datasets but also on cross-domain datasets.

The structure of this article is as follows. In Sect. 2, an overview of the current trends in the literature is provided. Section 3 presents a detailed description of the 3D-CAE-SiamAtt. The dataset description is presented in Sect. 4.1, followed by the evaluation metrics in Sect. 4.2, and the experimental results and discussion in Sect. 4.4 in Sect. 4. A series of ablation experiments are presented in Sect. 4.3. Finally, Sect. 5 concludes the study.

2 Related work

As discussed above, there exist various supervised deep learning models for hyperspectral image (HSI) classification with very high classification accuracy. However, these models necessitate an extensive amount of sampled/labeled training data which is difficult to get in the HSI domain. The problem of training deep networks with limited samples can amplify the risk of overfitting. With limited samples, there is a higher likelihood that the model may memorize the training data rather than learn meaningful patterns. Consequently, it may struggle to accurately classify new, unseen hyperspectral images [14, 16]. Also with limited samples, the network may struggle to propagate gradients effectively. In other words, the network's parameters might not update properly, leading to slow convergence or even stagnation in the learning process. This can hinder the model's ability to capture complex patterns in hyperspectral images [18]. Hence, there has been a pressing requirement for the advancement in deep learning models which can work with remarkable classification performance in scenarios where limited or no samples are available. For this, numerous supervised and semi-supervised models based on deep learning have been introduced for the classification of hyperspectral images (HSI), which are briefly reviewed as follows.

In 2016, one of the most noteworthy works was introduced by [19], who introduce a supervised three-dimensional convolutional neural network (3D-CNN) model specifically designed to tackle the challenges associated with limited data availability and class imbalance in HSI classification. The work also employs dropout and L2 regularization to mitigate the overfitting and vanishing gradients problem usually encountered in the erstwhile deep-learning models for HSI classification. In the similar

vein, Xu et al. [20] discuss a spectral–spatial unified network (SSUN) having a unified objective function with the incorporation of multiscale long short-term memory (LSTM). The experimental findings presented in the paper underscored its effectiveness in achieving accurate HSI classification, albeit with a high computational cost due to the model complexity. To mitigate the issue of vanishing gradients, researchers proposed a spectral–spatial residual network (SSRN) [21], which leverages residual blocks to enable more effective gradient backpropagation. This approach exhibits excellent performance when tested on datasets within the same domain. However, the SSRN fails to maintain its performance when extended to cross-domain datasets. Li et al. [22] address this issue by presenting a deep cross-domain few-shot learning model (DCFSL). They demonstrate that leveraging few-shot learning (FSL) techniques greatly aids in transferring knowledge from one domain to another. More specifically, DCFSL learns on one domain of datasets and extends this learning to unseen domains by following the meta-learning approach. Thus, DCFSL significantly enhances classification capabilities in scenarios where labeled samples are scarce. Afterward, many FSL-based deep learning models [15, 16, 22–24] for HSI classification have been introduced to further enhance the performance with limited labeled samples.

Among the aforementioned FSL-based HSI classifiers, the Siamese network introduced by [15] has gained more attention because of its ability to reduce the involved network parameters and reducing the overfitting problem. Therefore, various variations of the Siamese network such as [23–27], each with its own handcrafted characteristics, have been introduced, for HSI classification. One of them is proposed by [25] to introduce the Siamese-Based Bag-of-Features (S3BoF) network for improving feature extraction and reducing computational overhead by optimizing negative sample pairs. While the network achieves high overall accuracy, it exhibits slightly poorer performance on categories with a lower number of samples. To address this issue, [23] replace three-dimensional layers with one-dimensional and two-dimensional layers resulting in a reduction in network complexity. However, Xue et al.'s spectral–spatial-siamese-based network (S3Net) suffers from overfitting due to a high number of parameters and limited labeled samples involved. Another notable study conducted by [24] improve the discriminative ability of Siamese using a differentiated scale restricted graph convolutional network (DSR-GCN), which in turn helps in increasing the classification accuracy. In 2020, [26] made a big leap in the direction of achieving high classification accuracy by introducing a dual path Siamese-based CNN (Dual-SCNN) combined with adversarial-based training. By employing this approach, the authors intend to augment the diversity and quantity of training samples, thereby improving the network's capability to generalize and perform well on new and unseen examples. In the similar vein as of [26, 27] propose a three-dimensional semi-supervised Siamese network integrated with an autoencoder (3DAES). By merging the autoencoder with the Siamese network and training it primarily on unlabeled samples, the model effectively addresses the problem of limited labeled samples, as the autoencoder, which was initially proposed by [30] for unsupervised learning, consists of an encoder and a decoder that aim to reconstruct input data accurately, helps in augmenting the data. More specifically, the encoder transforms the input data α into a compressed hidden representation ω, while the decoder decodes the hidden representation into output α'. Both the encoder and decoder have symmetric connections with similar patterns. In another work, [17] combine autoencoder trained on unsupervised samples, with a convolutional neural network (CNN). Thus, authors aim to leverage the unsupervised learning capability of autoencoders along with the powerful feature extraction capabilities of CNNs, to get a boost in classification accuracy. Further, the Mei et al.'s model aligns well with the nature of HSI data and demonstrates promising results. To further enhance the performance, [29] discuss a spectral–spatial capsule network (CapsNet) by employing CNN, having the ability to extract features using weight-sharing convolutional kernels. Afterwards, various extensions to autoencoders such as layer-wise training [31], parameter reduction using PCA [32] and combinations of autoencoders with other models to enhance their capabilities in hyperspectral imaging (HSI) applications, have been introduced. To briefly summarize the outcomes of the aforementioned models with their classification accuracy on different datasets, along with the current trends in the literature of HSI classification, is presented in Table 1.

3 Proposed work

In this section, a novel convolutional autoencoder-based Siamese attention network is designed specifically for hyperspectral image (HSI) classification. While antecedent works have individually delved into components like the Autoencoder (3D-CAE), Siamese network, and attention layers, our innovation lies in their unprecedented integration, specifically tailored for the unparalleled enhancement of cross-domain and same-domain HSI classification. The strategic integration of architectural elements, coupled with a customized tuned model incorporating advanced regularization and optimization techniques, leads to unparalleled classification performance.

Table 1 Current Trends in Literature

References	Model	Learning	Dataset	Training ratio	Overall accuracy	Average accuracy	Kappa coefficient
[19]	3D-CNN	Su	IP	0.17	97.56	99.23	97.02
			PU	0.1	99.54	99.66	99.41
			KSC	0.1	96.31	94.68	95.9
[20]	SSUN	Su	PU	0.1	99.46	99.62	99.26
			IP	0.17	98.4	99.27	98.14
			KSC	0.1	97.71	97.17	97.45
[21]	SSRN	Su	IP	0.3	99.19	98.93	99.07
			PU	0.2	99.79	99.66	99.72
			KSC	0.3	99.61	99.33	99.56
[22]	DCFSL	S-Su	PU	205	83.65	83.77	78.7
			IP	205	66.81	77.89	62.64
			SA	205	89.34	94.04	88.17
[23]	S3-Net	S-Su	SA	320	95.95	98.23	95.49
			IP	320	91.54	95.58	90.38
			PU	180	96.27	96.91	95.11
[24]	DSR-GCN	S-Su	HU	75	81.11	82.40	79.57
			SA	180	97.35	96.64	97.04
			IP	180	82.29	88.64	79.94
			PU	45	89.49	91.01	86.39
[25]	S3BoF	Su	SA	0.05	99.81	99.82	99.79
			IP	0.05	96.82	95.28	96.38
			PU	0.05	99.73	99.49	99.64
[26]	Dual-SCNN	Su	KSC	0.003	89.95	89.03	88.7
			PU	0.001	84.9	85.28	80.54
[27]	3DAES	S-Su	SA	160	90.39	94.41	89.00
			YRE	200	90.02	84.97	88.00
			PU	90	79.63	82.97	74.00
[28]	3DVSCNN	S-Su	PU	90	76.43	75.47	69.00
			SA	160	89.17	94.07	88.00
[29]	CapsNet	Su	PU	0.01	94.9	95.42	93.24
			SA	0.06	99.94	99.95	99.92
[16]	MDL4OW	S-Su	PU	90	76.55	81.42	70.00
			SA	160	82.44	90.57	81.00
			YRE	200	87.46	83.41	86.00
[17]	3DCAE	Su	IP	0.1	92.35	92.04	–
			PU	0.05	95.39	95.36	–
			SA	0.05	95.81	97.45	–

Our proposed hybrid model, featuring a distinctive combination of encoder and decoder layers with customized parameters, is crafted to optimize HSI classification performance, presenting an experimental and novel contribution. The 3D-CAE, a pivotal component, plays a crucial role in learning data representations using unsupervised data, contributing significantly to the overall effectiveness of the model. The semi-supervised nature of our model is a key strength, addressing the challenge of limited labelled data in hyperspectral image classification. This aspect is vital for real-world applications where obtaining labelled data can be resource-intensive. The Siamese network employs three-dimensional convolutions and enhances feature separability by reducing the intraclass distance and increasing interclass distance, leveraging the high-dimensional and information-rich nature of hyperspectral data. Embedded attention layers, incorporating both spectral and spatial aspects, efficiently learn from

three-dimensional hyperspectral data. These layers reduce duplicate information and emphasize only significant channels and pixels using a binary cross-entropy loss function, enhancing the network's performance. Our proposed semi-supervised spectral–spatial network's effectiveness is validated through comprehensive experimentation, demonstrating exceptional performance in HSI classification tasks.

Figure 1 illustrates the overall structure of the proposed model that consists of a convolutional autoencoder embedded with Siamese action and spectral–spatial attention layers. The handcrafted autoencoder is trained with unlabelled samples to learn the best possible representation as discussed in 3.1. Next, two branches of the trained autoencoder are created and incorporated to form a Siamese network as detailed in 3.2. Each of the branches takes a different HSI cube and gives its own reconstructed output which are concatenated together. The concatenated representation is then passed through attention layers to highlight relevant features. Next, the output of the attention layers is used for multi-class classification as discussed in detail in Sect. 3.3. At last to summarize the working of the proposed network, Algorithm 2 is provided in Sect. 3.4.

3.1 Three-dimensional convolutional autoencoder (3D-CAE)

The proposed architecture for HSI classification consists of a three-dimensional convolutional autoencoder (3D-CAE) along with a Siamese network for similarity learning. Figure 2 illustrates the structure of the proposed autoencoder, and Table 2 provides an overview of its associated parameters. The overall structure of the model has an encoder and decoder, comprising a convolutional autoencoder. The attention layers are added after the decoder to capture relevant information and enhance the representation learned by the encoder. The attention layers help the autoencoder focus on important features or regions in the input data, improving its ability to reconstruct the input accurately. The encoder takes a hyperspectral cube, say $H_C \in \mathbb{R}^{(x\times y\times z)}$ where x and y represent the height and width of the cube, respectively, and z represents the number of spectral bands contained in the cube, the encoder component g_e of the autoencoder extracts discriminative features and produces an encoded feature representation $e_\phi^{H_C} \in \mathbb{R}^d$. The decoder component of the autoencoder then takes $e_\phi^{H_C}$ reconstructs the hyperspectral image $H_C^{'} \in \mathbb{R}^{(x\times y\times z)}$. During training with unlabeled samples, the autoencoder aims to reduce the loss between the cube inputted and cube reconstructed using binary cross entropy and improves convergence using the Adam optimizer.

The reconstructed outputs of the autoencoder are represented by Eq. 1, given as follows:

$$H_C^{'} = g_d(g_e(H_C)) \tag{1}$$

where g_e refers to the encoding function, which takes an input and maps it to a latent representation or feature vector. It is tasked with extracting distinctive features from the input data. g_d refers to the decoding function, which takes the latent representation produced by the encoding function and reconstructs the input data. It performs the mapping of the latent representation back to the original input space, with the objective of generating a reconstructed version of the input data. The convolutional layer having kernel Ke in Eq. 2

$$Ke \in \alpha^{x_{ke}\times y_{ke}\times z_{ke}\times ke_{\{in\}}\times ke_{\{out\}}} \tag{2}$$

where x_{ke}, y_{ke} and z_{ke} represent the height, width, and depth of the kernel, respectively, having ke_{in} as the input and ke_{out} as the output bands. The value of Ke impacts the network performance greatly. If the value of Ke is too high, the fine-grained subordinate features remain ignored. However, keeping the value of Ke too small increases the computational time due to the slow processing of the HSI cube. After the initial step of dimensionality reduction using principal component analysis and a window size of 9, the pre-processed data is divided into smaller cubes, and

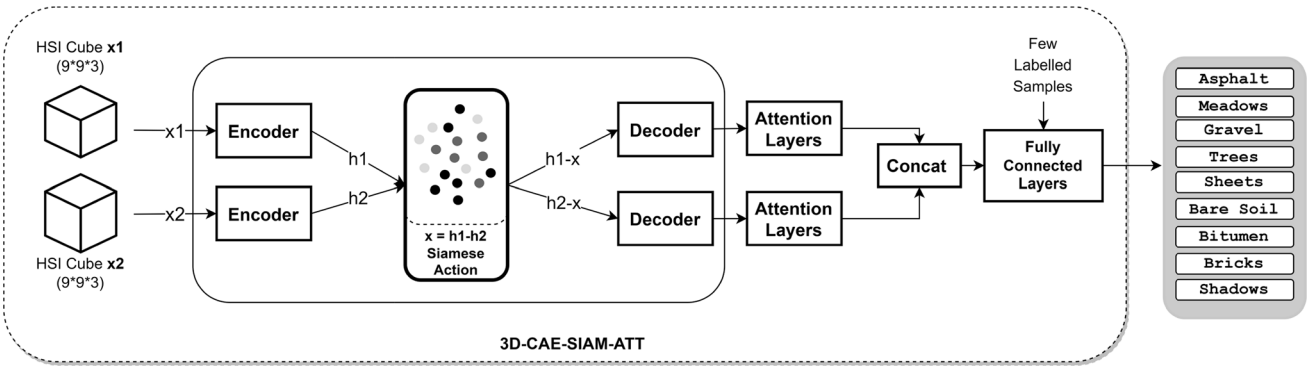


Fig. 1 Brief diagram of the proposed network

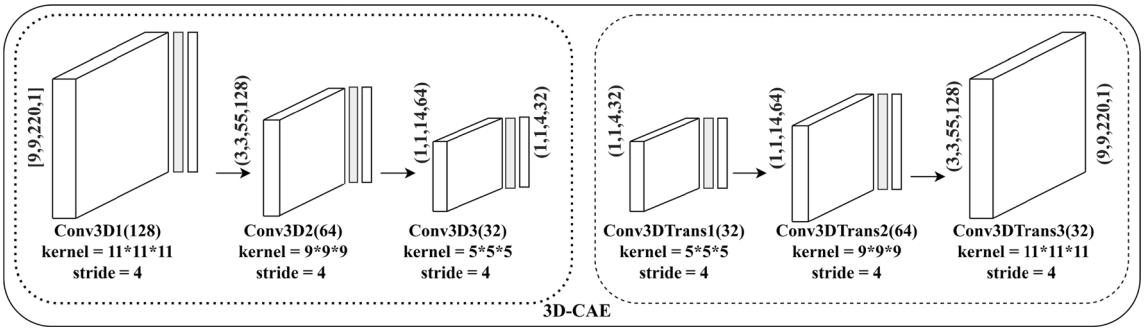


Fig. 2 3D-CAE

Table 2 Parameters of 3D-CAE

Layers	Type	Kernel number	Kernel size	Stride	Padding
CONV1	Conv3D + BN + L2 (0.1) + Activation(ReLU)	*N*	11×11×11	4×4×4	Same
CONV2	Conv3D + BN + L2 (0.1) + Activation(ReLU)	*N*	9×9×9	4×4×4	Same
CONV3	Conv3D + BN + L2 (0.1) + Activation(ReLU)	*N*	5×5×5	4×4×4	Same
DECONV1	Conv3DTranspose + L2 (0.1) + BN + Activation(ReLU)	*N*	5×5×5	4×4×4	Same
DECONV2	Conv3DTranspose + L2 (0.1) + BN + Activation(ReLU)	*N*	9×9×9	4×4×4	Same
DECONV3	Conv3DTranspose	*N*	11×11×11	4×4×4	Same
FC	Fully Connected + Sigmoid	2	–	–	–

padding is applied as needed. The resulting input tensor is then passed through a handcrafted autoencoder. This autoencoder is composed of three-dimensional convolutional layers with filter sizes 128, 64, and 32, respectively, each followed by batch normalization and ReLU activation. The inclusion of the batch normalization layer helps normalize the activations, facilitate gradient propagation, and stabilize the training process, while ReLU introduces nonlinearity to the model. The Conv3DTranspose layers in the decoder aid in learning the inverse mapping, allowing the network to generate meaningful and high-fidelity representations of the input hyperspectral data. The choice of kernel sizes, strides, and padding in the Conv3DTranspose layers is optimized to mitigate potential reconstruction issues. To enhance robustness and avoid overfitting, there are regularization techniques, including L2 regularization with a coefficient of 0.1, and batch normalization after each Conv3DTranspose layer. These measures are aimed at promoting stable training and reducing the risk of reconstruction-related artefacts.

The results of the reconstructed cubes are measured using Eq. 3, where L signifies loss. The similarity between the reconstructed and input cubes becomes higher as the value of L decreases. The autoencoder was trained using the Adaptive Moment Estimation optimizer, with binary cross entropy serving as the loss function.

$$L = ||H_{\mathrm{C}}^{'} - H_{\mathrm{C}}||^2 \tag{3}$$

3.2 Siamese-action (SIAM)

The Siamese network is created by combining two instances of the autoencoder, which share identical weights. The Siamese network takes pairs of inputs and learns to predict their similarity. The proposed inculcated Siamese network, depicted in Fig. 3 takes two parallel inputs in sample pairs. The encoders take a separate HSI cube x_1 and x_2 sliced from input HSI and generates its own embedding $e_{\phi}^{x_1}$ and $e_{\phi}^{x_2}$ by mapping it into a lower dimensional metric space. Next, the Euclidean distance $D(x_1, x_2)$ in the spectral space and spatial dimension is calculated using Eq. 4.

$$D(x_1, x_2) = ||(e_{\phi}^{x_1} - e_{\phi}^{x_2})||_2 \tag{4}$$

Thereafter, the embedding are modified as $e_{\phi}^{x_1} - D(x_1, x_2)$ and $e_{\phi}^{x_2} - D(x_1, x_2)$ and passed to embedded decoders for reconstruction. The more identical the two samples are, the more presumably they correspond to the same class. The loss function for the proposed Siamese is binary cross entropy summarized in the following Eq. 5:

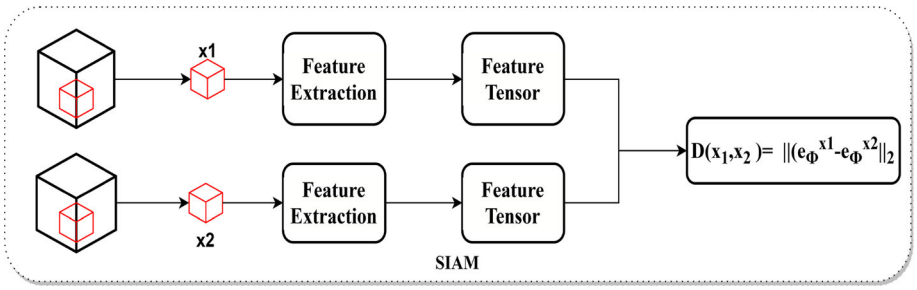


Fig. 3 Siamese-action

Algorithm 1 GenRandomBatch(inGroups, BatchHalfsize)

Initialize empty lists: OutImgA, OutImgB, OutScore
Create a list allGroups containing indices for all input data groups
for each matchGroup in [True, False] **do**
  Randomly choose batchHalfsize indices from allGroups and assign them to groupIdx
  for each cIdx in groupIdx **do**
    Randomly choose a sample from the corresponding group in inGroups and add it to OutImgA
  end for
  if matchGroup **then**
    Set bGroupIdx to be the same as groupIdx
    Append 1 to OutScore batchHalfsize times
  else
    Create an empty list nonGroupIdx
    for each cIdx in groupIdx **do**
      Randomly choose a different group index from allGroups and add it to nonGroupIdx
    end for
    Set bGroupIdx to be nonGroupIdx
    Append 0 to OutScore batchHalfsize times
  end if
  for each cIdx in bGroupIdx **do**
    Randomly choose a sample from the corresponding group in inGroups and add it to OutImgB
  end for
end for
Stack OutImgA, OutImgB, and OutScore along first dimension and return result

$$L_{bce} = -gt_n \log(\sigma(\gamma_n)) - (1 - gt_n) \log(\sigma(1 - \gamma_n)) \quad (5)$$

In this context, γ_n denotes the predicted probability of pixel n belonging to the other class, while $1 - \gamma_n$ represents the probability of the pixel belonging to the same class. The activation function σ is utilized as the sigmoid function, which compresses the predicted values between 0 and 1, interpreting them as probabilities. The logarithmic terms penalize deviations from the true labels. If the predicted probability aligns with the true label, the corresponding term approaches zero, contributing minimally to the overall loss. However, significant deviations lead to higher losses. The term $-gt_n \cdot \log(\sigma(\gamma_n))$ penalizes entries where the predicted probability $\sigma(\gamma_n)$ for samples belonging to the same class is close to zero. The network enhances separability between different classes by assigning low probabilities to dissimilar samples. The term $-(1 - gt_n) \log(\sigma(1 - \gamma_n))$ penalizes entries where the predicted probability for pixels belonging to the other class is close to zero. This encourages the network to assign higher probabilities to similar samples, bringing samples from the same class together. By optimizing the binary cross-entropy loss function, the Siamese network effectively guides the model to achieve enhanced separability, ensuring that the learned representations in the feature space align with the spatial and spectral characteristics of the HSI data. Notably, it influences the convergence speed, emphasizing faster convergence due to its ability to guide the network with clearer gradients. Additionally, the quality of learned representations is significantly improved, enabling the network to discern and highlight relevant features in hyperspectral data (Fig. 4).

3.3 Attention layers (ATT)

The attention layers were added to suppress the effect of noisy pixels [33]. The involved self-attention mechanism in the attention module extracts the significant pixels by exploiting the association between the hyperspectral pixels contained in the HSI cubes. This can be likened to the human eye's capability to focus on the essential elements in an input image. For the HSI cube $H_C \in \alpha^{x \times y \times z}$ having $x \times y$ as the spatial size and z as the spectral channels, the operation of the attention module is symbolized in Eqs. 6 and 7 where $M_{Spe} \in \alpha^{1 \times 1 \times C}$ denoting one-dimensional spectral map and $M_{Spa} \in \alpha^{x \times y \times 1}$ denotes two-dimensional spatial attention map [34]. After passing through the attention layers, labeled samples are added to the fully connected layer to rectify the unsupervised samples and train the network to perform the classification process. Figure 5 shows the spectral and spatial attention layers involved. Lastly, the addition of a flattening layer is essential for converting the output of the preceding layer, which is multidimensional, into a one-dimensional vector. This conversion is necessary when transitioning from convolutional layers to fully connected layers. The final layer of the model consists of a single output neuron which activates with the sigmoid function. For optimization during training, the popular Adam optimizer is utilized, which dynamically adjusts the learning rate to facilitate convergence.

$$H_C^{'} = M_{spe}(H_C) \otimes H_C \tag{6}$$

$$H_C^{''} = M_{spa}(H_C^{'}) \otimes H_C^{'} \tag{7}$$

3.4 3DCAE-SIAM-ATT algorithm

To summarize the working of proposed 3DCAE-SIAM-ATT model, an algorithm 2 which follows a multi-step process is provided below.

Algorithm 2 Pseudocode for 3DCAE-SIAM-ATT

Training Set $T = \{T_1, T_2, \ldots, T_n\}$; learning rate α, epochs e
Initialize weights and bias $W^i, b^i \ \forall i = 1$ to n
Build **3D-CAE()** inculcated with **SIAM()**
Build **ATT()**
for each iteration **do**
  for each $i \in [1, n]$ **do**
    Forward propagation
  end for
  for each $i \in [1, n]$ **do**
    Fine-tune W^i, b^i
  end for
end for

3D-CAE()
  Generate encoded tensor e_t and reconstructed tensor d_t using layers given in 2
  Minimize reconstruction error given in Equation 3
SIAM()
  Increase $InterCla_D$ and Reduce $IntraCla_D$
  Extract Patches from Training Set x_{i1}, x_{i2} using Algo 1
  Feed x_{i1}, x_{i2} to extract features y_{i1}, y_{i2}
  Calculate Distance **D** using Equation 4
  Train the network using L_{BCE} mentioned in 5 with **Adam** optimizer
ATT(i,n,a)
  For input tensor i, spectral channels n, attention units a
  a_w = dense(i,a)
  r_w = reshape(a_w,[1,1,1,n])
  a_i = ElementWiseMultiply(i, r_w)
  o_t = sum(a_i)
  Return o_t

return $W^i, b^i \ \forall i = 1$ to n

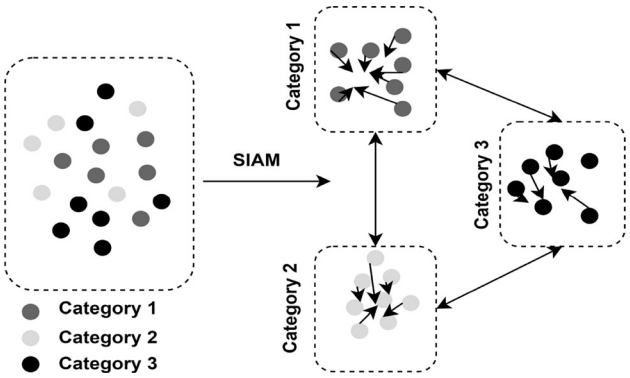


Fig. 4 Siamese increasing the inter-class and decreasing intra-class distance

The algorithm establishes an autoencoder that comprises an encoder and a decoder. The encoder utilizes CNN architectures to extract significant features from the input images. These extracted features are subsequently employed by the decoder to reconstruct the original images. The autoencoder is trained using the available training data, enhancing its performance through the Adam optimizer and minimizing the binary cross-entropy loss. The training process involves specifying parameters such as the number of epochs, batch size, and learning rate. Once the autoencoder is trained, a similarity model is constructed on top of it. This model includes a dense layer with sigmoid activation, which predicts the similarity between pairs of images based on their encoded features. The similarity model is then trained using a generator that produces batches of image pairs and their corresponding similarity labels. The training progress is tracked, and metrics such as loss, accuracy, recall, and precision are recorded. It is to be noted that an attention mechanism is implemented, which highlights salient regions in the reconstructed images, to enhance the reconstruction process.

4 Experimental results and discussion

The experimental results and discussion section of this paper is divided into four sub-sections. Firstly, a comprehensive description of the datasets utilized in this study is provided in the first Sect. 4.1. Section 4.2 provides an overview of the evaluation metrics employed to evaluate the performance of the proposed model in comparison to state-of-the-art (SOTA) models. Furthermore, Sect. 4.4 presents the experimental results, including a comparative analysis with the SOTA models showcasing the performance of the proposed model, while delving into ablation studies in Sect. 4.3 which are conducted during the experimentation process.

4.1 Dataset description

To assess the performance of 3D-CAE-Siam-Att, four well-known HSI datasets, namely Indian Pines (IP), Salinas (SA), Pavia University (PU) and Pavia Center (PC), were utilized. The details about each of the aforementioned dataset's HSI collections and their properties are provided in the following sub-sections. Figure 6 shows the sample distribution for these benchmark HSI datasets.

4.1.1 Indian Pines

The IP dataset under consideration exhibits a spatial resolution of 145 pixels and encompasses 224 spectral reflectance bands, spanning a wavelength range of 0.43 to 2.5 μm. This dataset is spread over 10,249 samples and not mutually exclusive of sixteen categories. After excluding the bands affected by water absorption, the dataset is reduced to a final count of 200 spectral bands. Further, the IP dataset consists of 10 categories as depicted in Fig. 6a along with their sample distribution. Additionally, Fig. 7a illustrates the composite of Indian Pines, while Fig. 7b showcases the corresponding ground truth.

4.1.2 Pavia University

The PU dataset has a height and width of 610 and 340 pixels and a spatial resolution of 1.3m per pixel. Each pixel in the Pavia University (PU) dataset consists of 115 spectral channels with wavelengths ranging from 0.43 to 0.86 μm. The final count of spectral channels is 103, after removing 12 noise-polluted bands. Further, this dataset has 42,776 samples dispersed in nine categories as depicted in

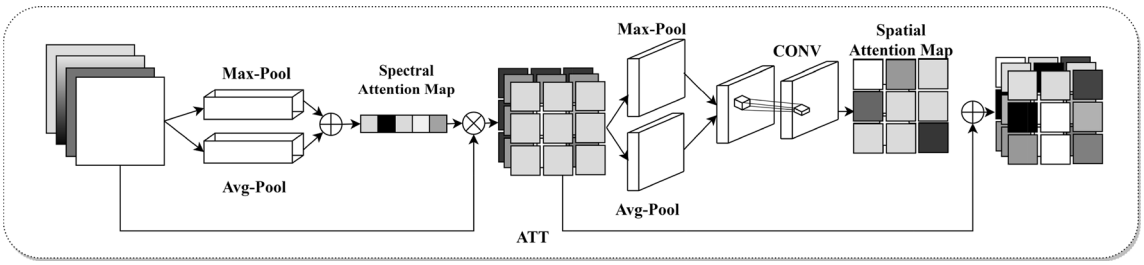


Fig. 5 Attention network

Fig. 6 Sample Distribution in (**a**) Indian Pines (**b**) Salinas (**c**) Pavia Center (**d**) Pavia University

Label	Count
Alfafa	46
CornNotill	1428
Cornmintill	830
Corn	237
GrassPast	483
GrassTrees	730
GrassPastM	28
HayWindrowed	478
Oats	20
SoybeanNotill	972
SoybeanMintill	2455
SoybeanClean	593
Wheat	205
Woods	1265
BuildingsGTD	386
StoneST	93

(a) Indian Pines

Label	Count
Brocoli1	2009
Brocoli2	3726
Fallow	1976
FallowRough	1394
FallowSmooth	2678
Stubble	3959
Celery	3579
GrapesUn	11271
SoilV	6203
CornGreenWeeds	3278
Lettuce4wK	1068
Lettuce5wK	1927
Lettuce6wK	916
Lettuce7wK	1070
VinyardU	7268
VinyardVertical	1807

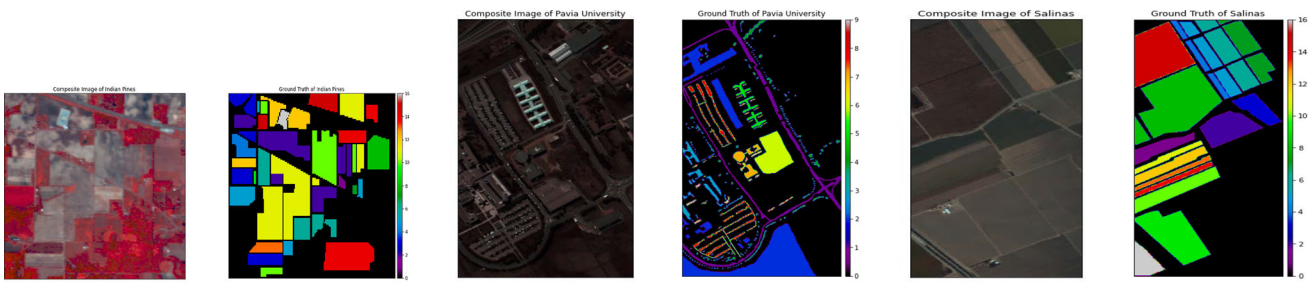
(b) Salinas

Label	Count
Water	824
Trees	820
Asphalt	816
SB Bricks	808
Bitumen	808
Tiles	1260
Shadows	476
Meadows	824
BareSoil	820

(c) Pavia Center

Label	Count
Asphalt	6631
Meadows	18649
Gravel	2099
Trees	3064
PMSheets	1345
BareSoil	5029
Bitumen	1330
SelfBB	3682
Shadows	947

(d) Pavia University



(a) IP-C (b) IP-GT (c) PU-C (d) PU-GT (e) SA-C (f) SA-GT

Fig. 7 Composite (C) and Ground Truth (GT) of Indian Pines (IP), Pavia University (PU) and Salinas (SA)

Fig. 6d. Additionally, Fig. 7c illustrates the composite of Indian Pines, while Fig. 7d showcases the corresponding ground truth.

4.1.3 Salinas

The SA dataset consists of images with a width of 512 pixels and a height of 217 pixels. It includes 224 spectral bands spanning the wavelength range of 0.43 to 2.5 μm.

After excluding the noisy bands corresponding to water absorption, the final dataset comprises 204 bands. The SA is spread into 54,129 samples and sixteen categories, which are depicted in 6b with their sample distribution. Figure 7e and Fig. 7f depicts the composite and ground truth of Salinas, respectively.

4.1.4 Pavia Center

The PC dataset comprises images with a height of 1096 pixels and a width of 715 pixels, offering a spatial resolution of 1.3m per pixel. Within each pixel of the dataset, there are 115 spectral channels, spanning widths ranging from 0.43 to 0.86 μm. However, the final count of spectral channels is 102 after removing 13 noise-polluted bands. Further, the PC dataset has 42,776 samples dispersed in nine categories, as depicted in Fig. 6c.

4.2 Evaluation Metrics

To evaluate and comparatively analyze the performance of the proposed model with the SOTA models, three commonly used parameters, which include Overall Accuracy (OA), Average Accuracy (AA), and Kappa Coefficient (KC), have been considered as the evaluation metrics. The OA metric quantifies the percentage of accurately classified pixels within the complete HSI dataset, such that $OA = \frac{C'}{C}$, where C' is the total number of samples classified correctly. Therefore, the OA offers a comprehensive evaluation of the model's classification performance across all the classes and can be computed using the following equation:

$$\mathrm{OA} = \frac{(\mathrm{TP} + \mathrm{TN})}{(\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN})} \tag{8}$$

where TP, TN, FP, and FN represent the abbreviations for true positives, true negatives, false positives, and false negatives, respectively. In contrast, the Average accuracy (AA), also known as mean class accuracy, determines the average classification accuracy for each specific class within the HSI dataset. By calculating the average accuracy per class, this metric offers valuable insights into the model's performance at a class-specific level, enabling a more detailed evaluation of its classification capabilities. The Average accuracy (AA) can be computed using Eq. 9, detailed as follows:

$$\mathrm{AA} = \frac{A_1 + A_2 + A_3 + \cdots + A_n}{n} \tag{9}$$

The kappa coefficient (KC), also known as Cohen's kappa, is a statistical metric used to evaluate the agreement between predicted and true class labels while considering the possibility of the agreement by chance alone. This measure takes into account both the classification accuracy and the potential agreement that could occur randomly and can be computed using Eq. 10, given below.

$$\mathrm{KC} = \frac{\mathrm{OA} - P_c}{1 - P_c} \tag{10}$$

where P_c denotes the speculative probability of possibility and can be calculated using Eq. 11:

$$P_c = \frac{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN}) + (\mathrm{FN} + \mathrm{TN})(\mathrm{FP} + \mathrm{TN}))}{(\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN})^2} \tag{11}$$

4.3 Ablation Studies

To study the impact of change in different parameters and hyper-parameters and other components of the proposed model, ablation studies have been conducted, by considering the IP dataset. This study helps in finding the best setting of the model along with the architecture for optimal performance and is organized into six sub-sections detailed as follows. Section 4.3.1 examines the influence of the training ratio on classification performance, Sect. 4.3.2 explores the impact of different kernel sizes, Sect. 4.3.3 investigates the consequences of removing attention layers, Lastly Sect. 4.3.4 delves into the implications of utilizing different optimizers.

4.3.1 Percentage of Training Samples

The findings in Fig. 8a demonstrate the influence of varying amounts of training sample sizes on the performance of the proposed model, as measured by OA, AA, and KC. It can be observed that when the network is trained with 40% unsupervised samples, the OA is relatively low. However, as the sample size increases to 60%, there is a noticeable improvement in performance. A significant gain in performance is observed when the training samples reach 80%. However, the overall accuracy is only marginal when the training samples are increased to 90%. These findings suggest that the proposed model achieves the best performance when trained with 80% unsupervised samples. The training performance and classification maps provided in Fig. 9 serve to reinforce these findings by demonstrating a decrease in misclassifications with an increase in training samples. The model exhibits excellent convergence, signifying its ability to learn and enhance performance. Furthermore, there is minimal or no overfitting, and the model demonstrates strong generalization capabilities, accurately performing on both the training and validation datasets.

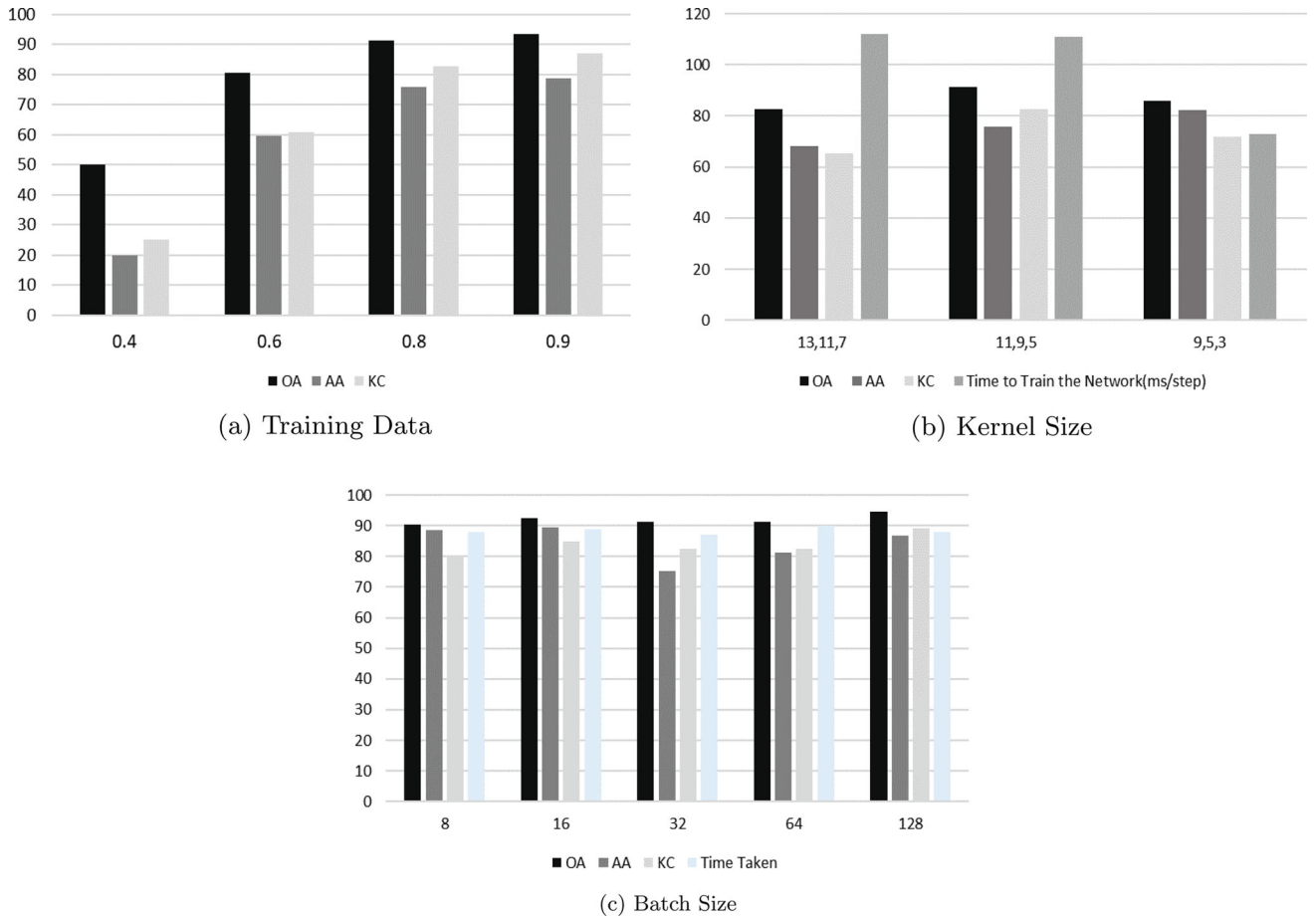


Fig. 8 Impact of different parameters

4.3.2 Impact of Kernel Size

The performance of the network has also been assessed by varying the kernel sizes used in the autoencoder, as presented in Fig. 8b. The results show that the best overall accuracy was obtained with kernel sizes 11, 9, and 7. Furthermore, Fig. 10 highlights that the proposed model achieves the most favorable convergence, as depicted in Fig. 10b, and exhibits the lowest misclassifications, as shown in Fig. 10e, when the autoencoder employs kernel sizes 11, 9, and 5.

4.3.3 Removal of Attention Layers

To assess the impact of the integrated attention layers on the model's performance, a comprehensive analysis was conducted. The classification performance was systematically evaluated by removing these attention layers, and the outcomes are illustrated in Fig. 11. The results demonstrate a notable decrease in overall accuracy (OA), average accuracy (AA), and Kappa coefficient (KC) for the model without attention. Specifically, the OA, AA, and KC metrics dropped from 91%, 75.83% and 82.6% dropped to 82.6%, 51.85%, and 65.21%, respectively. Interestingly, despite the decrease in accuracy, the training process exhibited a significant speed improvement, with the network achieving a training time of only 51 ms/step after the attention module's removal, compared to 111 ms/step with the attention layers. The classification accuracy while training and classification map is included in Fig. 11.

This observed reduction in training time without attention layers can be attributed to the streamlined computational process. Attention layers introduce additional computational overhead due to their sequential processing and recalibration operations. When removed, the model's training process becomes more efficient and faster, allowing for quicker convergence during the training phase. However, this efficiency comes at the expense of a reduction in classification accuracy, as the attention layers play a crucial role in capturing and emphasizing relevant features for improved classification performance.

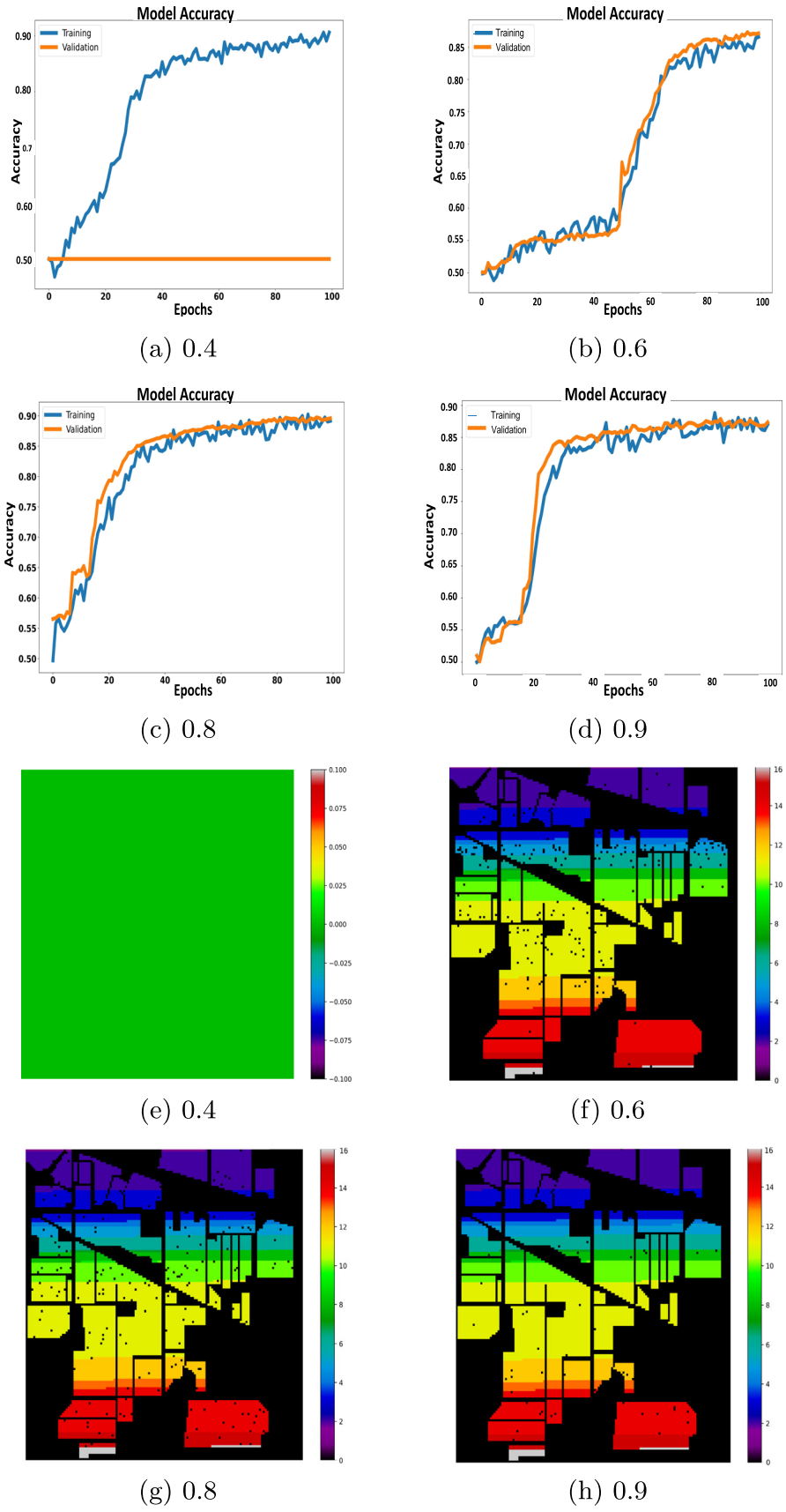


Fig. 9 Classification accuracy and classification maps for varying values of n

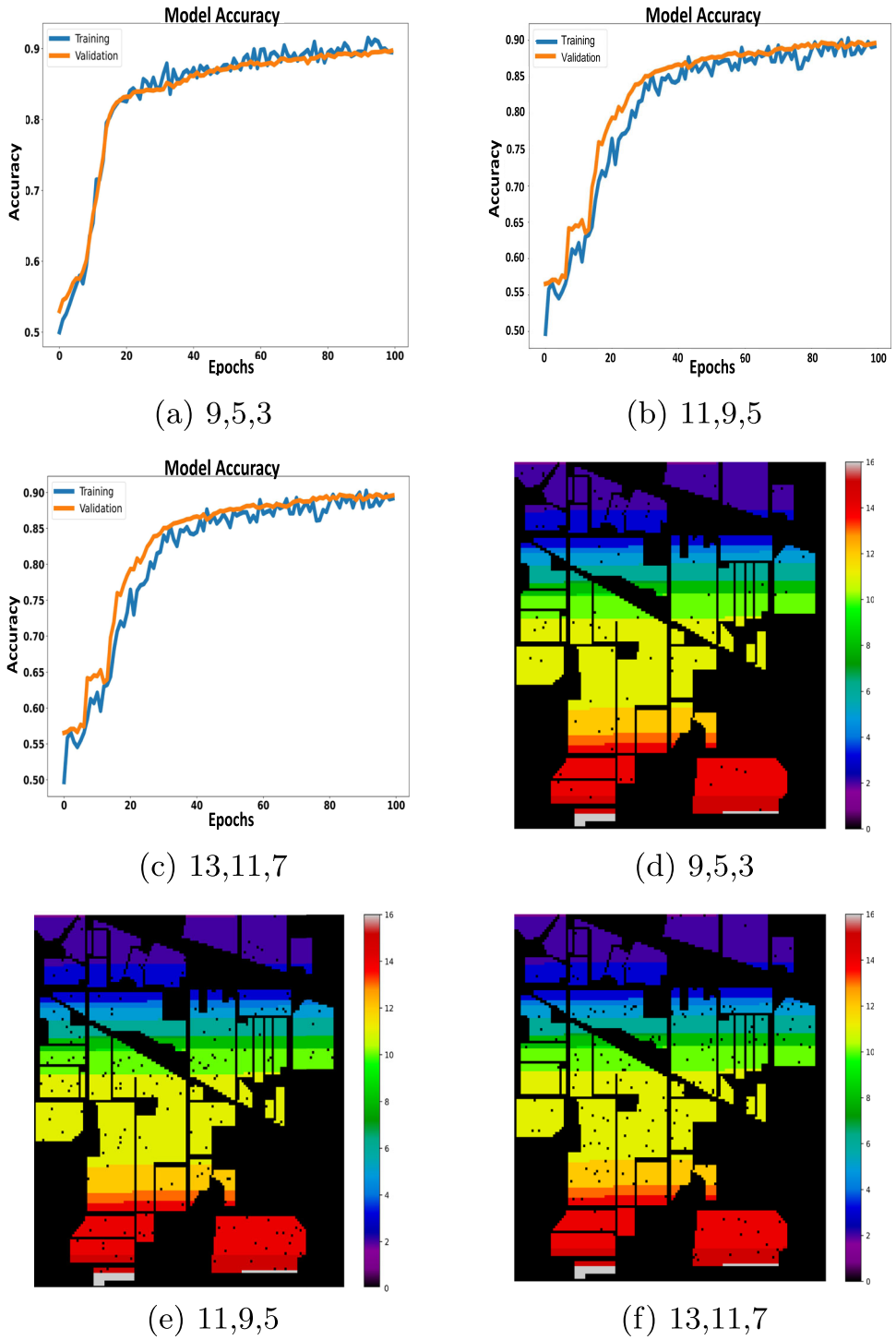


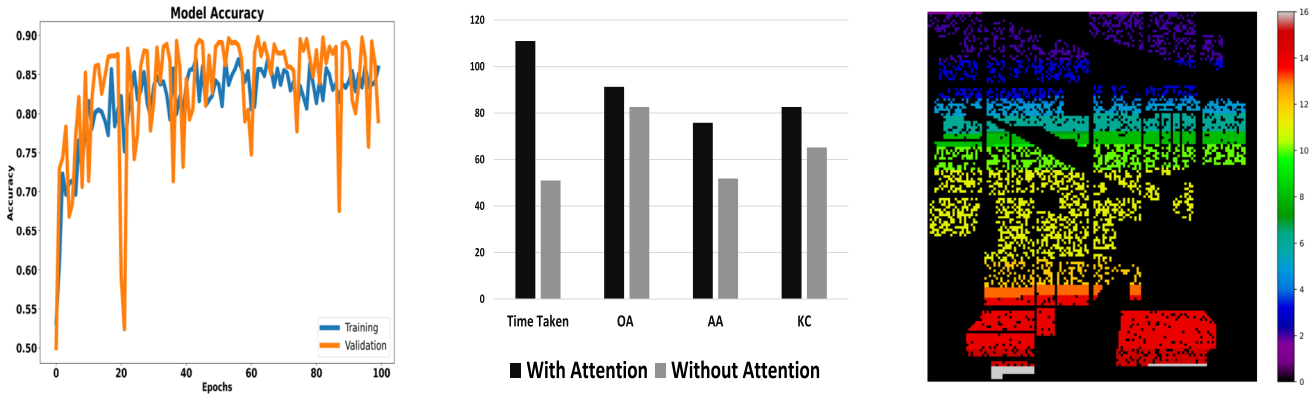
Fig. 10 Classification accuracy and classification maps for varying sizes of Kernel

4.3.4 Optimizers

To explore the performance of the proposed model with different optimizers, the three most commonly used optimizers were chosen for experimentation. The results of this experimentation are presented in Fig. 12, which clearly indicates that the accuracy of the proposed model is the lowest when optimized with RMSprop, better when optimized with AdaGrad, and the best when optimized with the Adam optimizer.

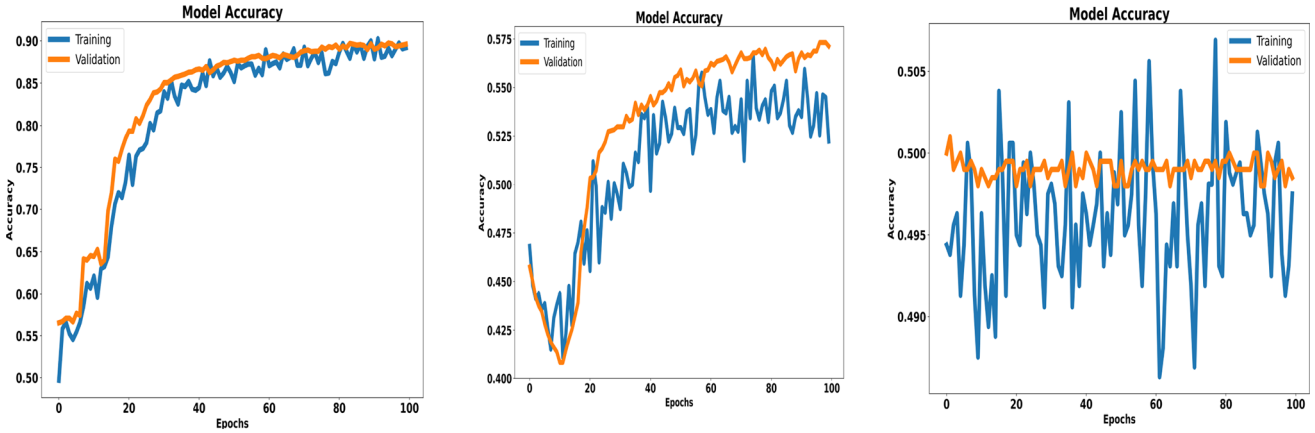
4.4 Comparison with other models

To comparatively analyze the performance of the 3D-CAE-SiamAtt model, several SOTA models such as Dual-SCNN- AT-Mixup [26], Dual-SCNN-Mixup [26], CapsNet



(a) Classification Accuracy (b) Time Taken and OA (c) Classification Map

Fig. 11 Performance evaluation after removal of attention layers



(a) Adam (b) Adagrad (c) RMSprop

Fig. 12 Effect of different optimizers on accuracy

[29], 3D-CNN [19], SSUN [20], SSRN [21], DCFSL [22] and MDL4OW [27], have been considered and performance has been evaluated on all the aforementioned datasets. Firstly the comparison is done in terms of OA, AA, and KC by detailing the experimental finding in Tables 3, 4 and 5, for the IP, PU and SA datasets, respectively, where the best performances are highlighted in bold. The tables clearly demonstrate that the proposed 3D-CAE-SiamAtt model achieves an overall accuracy of 91.30 on the IP dataset, 87.24 on the PU dataset, and 93.60 on the SA dataset.

In other words, the 3D-CAE-SiamAtt achieves the highest overall accuracy for IP, second highest for PU and third highest for SA dataset. Other models that achieved higher accuracy included SSRN for PU, and SSRN, 3DAES and 3DLSN for SA. The performance of SSRN is better due to the supervised nature of this study which trains the network using labeled data. Although semi-supervised models have the advantage of leveraging a larger dataset for training, the absence of explicit labels for the majority of the data introduces additional uncertainty and can make the learning process more challenging. As a result, the classification accuracy of semi-supervised models may not match the performance of supervised models. On other hand, 3DAES the semi-supervised model built using Siamese achieved better accuracy for SA, however it was marginally lower for other two datasets. It is to be noted that the proposed model achieves better or comparable results because of the involved three-dimensional convolution layers and attention layers which fit the nature of HSI considering both spatial and spectral information. Further. the involved attention layers discriminate the significant bands and ignore the noisy bands. Furthermore, it is to be noted that the proposed network took 111 ms/step for IP, PU and SA while 103 ms/step for PC with a total of 100 epochs having 50 steps each.

Secondly, the performance of 3D-CAE-SiamAtt has been comparatively analyzed by considering the cross-domain datasets, as the proposed model runs for not just same-domain but cross-domain datasets as well, provided

Table 3 $T_r > \text{IP}; T_e > \text{IP}$

	2D-CNN [25]	3D-CNN [25]	Hybrid SN [25]	3D-CNN [19] in [23]	SSUN [20] in [23]	SSRN [21] in [23]	DCFSL [22] in [23]	3DAES [27] in [23]	3D-CAE-SiamAtt
C1	77.67	71.42	91.78	98.55	**100**	89.36	**100**	97.83	**100**
C2	86.51	78.32	84.36	65.49	65.86	**87.66**	74.09	62.76	32.77
C3	84	70.23	82.41	79.63	81.12	**84.47**	81.23	69.83	49.51
C4	68.98	72.9	94.39	93.86	**96.73**	79.39	96.31	87.2	51.89
C5	**95.09**	87.18	94.14	89.33	90.43	95.05	91.84	87.7	68.73
C6	**97.95**	92.94	93.3	93.87	90.15	96.94	97	92.94	90
C7	71.15	86.78	88.58	99.55	98.57	63.68	**100**	**100**	**100**
C8	98.46	95.53	98.11	99.42	98.62	**99.6**	98.34	98.77	96.44
C9	90	61.11	71.69	98.75	**100**	43.56	**100**	**100**	**100**
C10	87.13	75.17	**89.2**	78.1	73.16	76.59	80.15	72.23	59.05
C11	87.77	77.9	**94.1**	63.34	71.43	89.97	72.31	62.15	41.09
C12	75.67	71.34	**86.16**	71.76	80.8	73.46	78.05	72.06	46.88
C13	97.42	96.43	96.49	98.2	99.08	90.6	99.68	**100**	99.51
C14	**98.32**	92.6	90.19	83.07	88.29	96.97	93.8	91.64	87.27
C15	85.79	80.63	85.44	90.62	**96.31**	83.59	94.29	82.44	90.15
C16	97.84	97.22	81.98	98.82	99.86	82.68	99.59	**100**	**100**
OA	89.09	81.3	89.57	77.13	80.05	87.09	83.1	75.52	**91.3**
AA	87.48	81.73	88.9	87.65	89.4	83.35	**91.04**	86.1	75.83
KC	**87.58**	78.62	88.12	74.25	77.44	85.35	80.87	72.32	82.6

Table 4 $T_r > PU; T_e > PU$

	Dual- SCNN- AT-Mixup [26]	Dual-SCNN-Mixup [26]	CapsNet [29] in [26]	3D-CNN [19] in [23]	SSUN [20] in [23]	SSRN [21] in [23]	DCFSL [22]	MDLOW [16] in [27]	3DAES [27]	3D-CAE-SiamAtt
C1	80.76	81.33	80.18	83.28	84.7	**91.79**	82.20	72.77	78.49	43.73
C2	**89.96**	89.9	82.26	85.96	81.68	84.51	87.74	76.24	76.8	73.01
C3	69.51	63.56	64.11	81.89	73.12	**85.23**	67.46	70.39	64.96	35.58
C4	96.24	96.93	89.8	90.57	**97.72**	94.1	93.16	93.21	92.91	89.84
C5	99.49	98.96	97.73	99.92	99.18	**99.99**	99.49	99.41	95.27	95.98
C6	77.6	79.27	81.85	84.93	83.37	**97.67**	77.32	71.44	82.19	85.06
C7	77.23	79.35	78.83	94.9	89.76	**97.72**	81.18	88.03	84.19	75.93
C8	80.6	74.57	81.27	64.4	**92.56**	91.77	66.73	63.8	79.13	47.82
C9	97.59	99.58	93.64	60.06	99.61	99.76	**98.66**	97.47	92.76	92.5
OA	85.81	85.41	81.9	83.84	85.2	**89.75**	83.65	81.42	82.97	**87.24**
AA	85.44	84.56	83.28	82.88	89.08	**93.62**	83.77	76.55	79.63	71.05
KC	81.36	81.01	76.65	79.02	80.92	**86.84**	78.70	70	74	74.49

that number of bands are similar. Table 6 lists the results achieved for cross-domain datasets, where four scenarios (a) training on IP, testing on SA; (b) training on SA, testing on IP; (c) training on PU, testing on PC; (d) training on PC, and testing on PU, have been considered. The provided experimental results clearly validates the proposed model's potential to train and test on cross-domains. More specifically, the proposed model achieves its best overall accuracy when training is performed on IP and testing on SA. However, it can also be observed that the cross-domain

Table 5 $T_r > SA; T_e > SA$

	SSFTT [24]	ASPCNet [24]	3DCNN [19] in [23]	SSUN [20] in [23]	SSRN [21] in [23]	3DLSN [28]	3DCAE [17] in [27]	MDL4OW [16] in [27]	3DAES [27]	3D-CAE-SiamAtt
C1	99.49	95.84	99.41	97.05	**100**	99.87	99.28	86.12	99.97	98.8
C2	99.84	**99.93**	99.46	95.95	99.64	99.46	59.04	91.97	99.18	97.74
C3	99.83	98.68	99.61	98.61	**99.89**	97.93	66.54	80.34	95.15	64.67
C4	95.84	91.88	99.17	**99.88**	97.81	99.75	98.65	99.28	96.47	98.56
C5	96.38	97.46	98	**98.54**	98.17	95.23	81.94	95.06	91.18	96.26
C6	98.28	98.64	99.83	99.72	**99.94**	99.75	98.52	99	99.17	93.81
C7	99.46	99.59	99.2	97.13	99.81	99.27	97.31	98.16	**99.9**	97.76
C8	75.6	86.07	78.79	80.54	88.15	75.05	68.11	47.38	75.66	**95.9**
C9	**99.97**	97.15	99.4	99.56	99.56	99.25	95.06	97.43	98.96	98.58
C10	89.01	93.62	92.89	95.6	**96.34**	92.92	92.43	87.36	89.74	84.28
C11	**99.91**	96.23	98.85	98.38	97.3	98.68	72.26	96.16	99.54	92.97
C12	99.26	96.28	99.75	99.77	99.04	99.86	72.16	99.76	99.03	**100**
C13	97.98	89.31	99.64	99.38	98.72	99.98	99.78	99.75	99.76	**100**
C14	96.13	94.17	**99.36**	98.6	95.09	98.89	89.93	98.1	97.24	96.35
C15	82.85	77.33	87.82	84.39	79.1	80.87	56.98	77.66	81.7	**96.21**
C16	98.27	**100**	96.85	97.08	98.93	94.65	44.35	95.64	87.94	96.01
OA	91.26	91.5	93.01	92.61	93.83	**95.71**	75.58	90.57	**94.41**	**93.6**
AA	95.51	94.51	**96.75**	96.26	96.72	91.05	72.64	82.44	90.39	94.24
KC	90.3	90.55	92.24	91.79	**93.13**	90	70	81	89	87.2

Table 6 Cross-domain datasets

	$T_r >$ IP; $T_e >$ SA 3D-CAE-SiamAtt	$T_r >$ SA; $T_e >$ IP 3D-CAE-SiamAtt	$T_r >$ PU; $T_e >$ PC 3D-CAE-SiamAtt	$T_r >$ PC; $T_e >$ PU 3D-CAE-SiamAtt
C1	100	65.21	97.9	51.54
C2	99.19	26.75	98.6	67.48
C3	99.79	29.51	88.99	69.31
C4	100	24.89	97.5	77.67
C5	89.39	33.54	79.14	56.87
C6	96.33	47.8	97.9	75.4
C7	97.59	75	79.67	57.81
C8	94.21	85.56	97.8	88.94
C9	100	90	98.81	50.47
C10	50	30.76	–	–
C11	100	34.54	–	–
C12	100	21.92	–	–
C13	92.13	47.8	–	–
C14	65.7	51.69	–	–
C15	96.34	42.48	–	–
C16	34.25	73.11	–	–
OA	87.2	73.91	78.91	75.32
AA	88.43	48.78	92.92	66.17
KC	74.41	47.82	57.82	50.64

results are on a slightly lower side than those achieved on the same-domain datasets. This could be because cross-domain datasets present unique challenges due to variations in distribution, domain-specific features, and limited

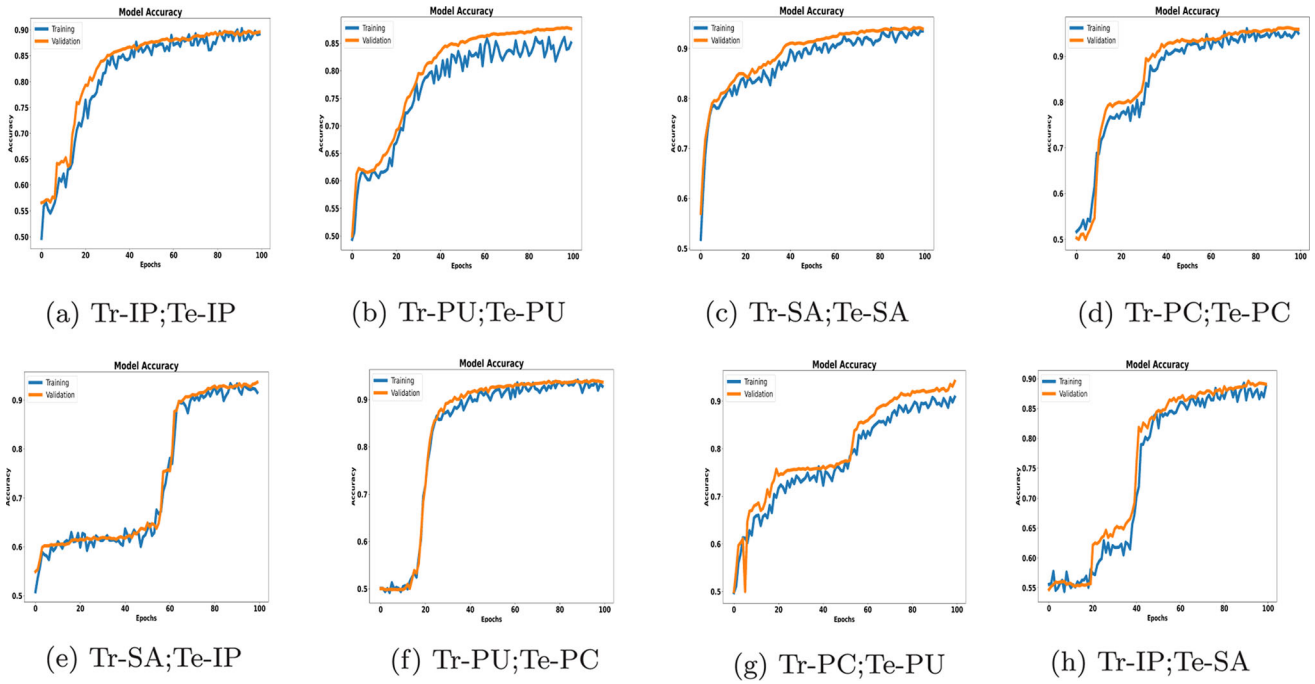


Fig. 13 Model accuracy graphs achieved

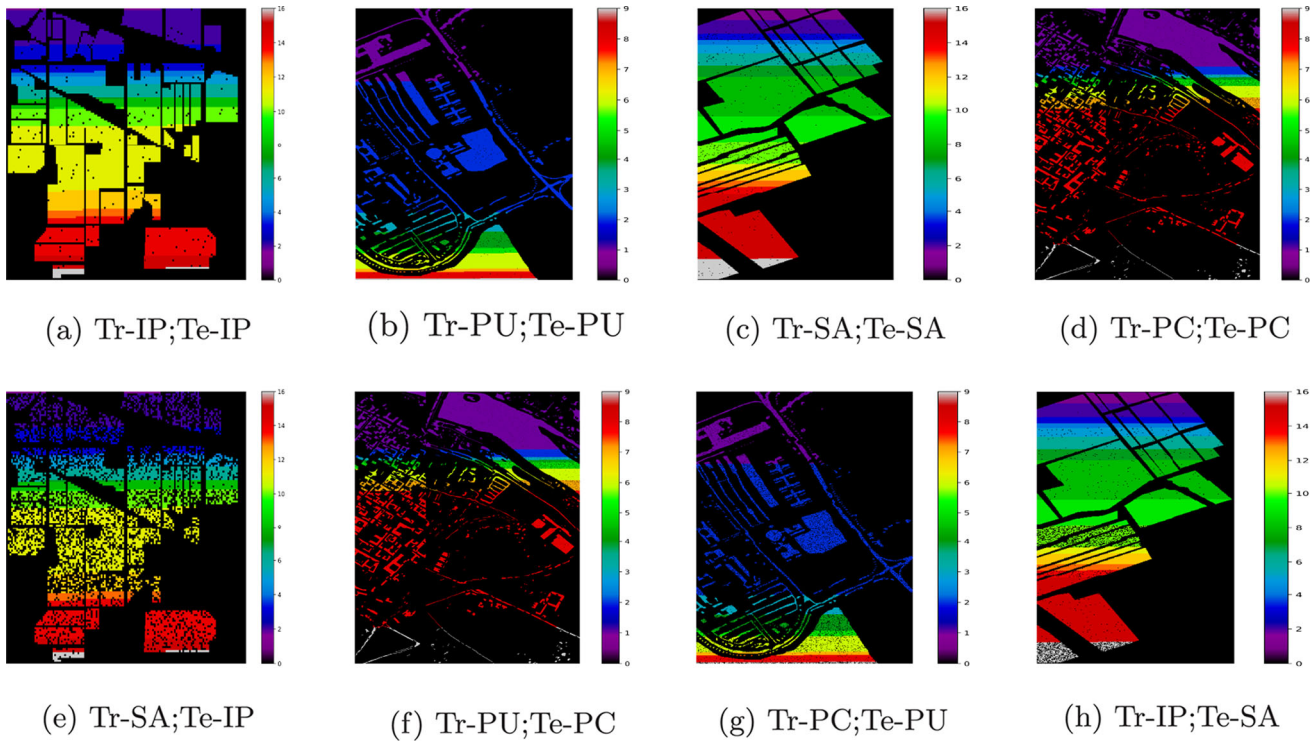


Fig. 14 Classification maps for same-domain and cross-domain datasets

labeled data availability compared to same-domain datasets. However, these challenges provide valuable opportunities for advancements in the field of cross-domain classification. By addressing domain shift through techniques such as domain adaptation and leveraging feature augmentation, future models can improve the performance of models on cross-domain datasets. Moreover, developing specialized architectures and exploring innovative approaches can further enhance the classification results. These endeavors will contribute to a deeper understanding

of cross-domain classification and drive progress in solving complex real-world problems spanning different domains.

To further qualitatively validate the performance of the model, the training accuracy graphs for the same domain and cross-domain datasets are demonstrated in Fig. 13 which showcases the classification maps when experimented on the same or cross-domain datasets. The presented classification maps evidently show the similarity of obtained maps to the ground truth images which also means that the ability of the model to work cross-domain and produce results is excellent. Further, the 3D-CAE-SiamAtt produces excellent classification maps with minimum noise when $T_r > \text{IP}; T_e > \text{IP}$, $T_r > \text{PU}; T_e > \text{PU}$ and $T_r > \text{SA}; T_e > \text{SA}$. Moreover, the accuracy graphs to depict the performance of the proposed network while training the model are presented in Fig. 13. On one hand, graphs trained for the same domain data produced smooth learning curves with no overfitting, there was slight overfitting observed while training PC and testing PU. All other datasets have smooth graphs validating the performance of the proposed model (Fig. 14).

5 Conclusion

This paper presents a novel semi-supervised deep learning network for HSI classification, which incorporates an autoencoder along with Siamese and attention layers. The convolutional autoencoder is trained in an unsupervised manner to refine the representation, solving the problem of limited samples. The handcrafted Siamese network enhances the feature separability between samples, by increasing the distance in samples belonging to different classes and minimizing the distance in samples belonging to the same class. The attention layers aided in extracting the informative spectral–spatial features by suppressing the noisy or less important ones. Four benchmark HSI datasets were used to validate the tenacity of the proposed 3D-CAE-SiamAtt. The proposed model garnered a classification accuracy of 91.3 and 93.6 for Indian Pines and Salinas, respectively, with a highest of 87.27 when tested on cross-domain data. The experiments validate the domain adaptation capability of 3D-CAE-SiamAtt. Thus, the potential applications of the proposed HSI classification model can extend beyond its current domains of remote sensing, environmental monitoring, agriculture, and mineral exploration. As a future work, the model can be optimized to enhance the convergence speed for effectively utilizing a broader range of fields and disciplines.

Author contributions Pallavi Ranjan conceived the idea, performed implementation, conducted experiments, and wrote the original manuscript. Rajeev Kumar performed writing—review and editing. Ashish Girdhar identified the problem and provided guidance throughout the project.

Funding No specific grant was received for this research from any public, commercial, or non-profit funding agency.

Data availability The hyperspectral datasets are open and available on https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

Code availability The code used for the experiments in this research project is available upon request via mail.

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

1. 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
2. Sellami A, Farah M, Farah IR, Solaiman B (2019) Hyperspectral imagery classification based on semi-supervised 3-d deep neural network and adaptive band selection. Expert Syst Appl 129:246–259
3. Ball JE, Anderson DT, Chan CS (2017) Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. J Appl Remote Sens 11(4):042609–042609
4. Nalepa J, Myller M, Imai Y, Honda K, Takeda T, Antoniak M (2020) Unsupervised segmentation of hyperspectral images using 3-d convolutional autoencoders. IEEE Geosci Remote Sens Lett 17(11):1948–1952
5. Garea AS, Heras DB, Argüello F (2019) Caffe CNN-based classification of hyperspectral images on GPU. J Supercomput 75:1065–1077
6. Zhou F, Hang R, Liu Q, Yuan X (2019) Hyperspectral image classification using spectral–spatial LSTMS. Neurocomputing 328:39–47
7. Li K, Wang M, Liu Y, Yu N, Lan W (2019) A novel method of hyperspectral data classification based on transfer learning and deep belief network. Appl Sci 9(7):1379
8. Zhong P, Gong Z, Li S, Schönlieb C-B (2017) Learning to diversify deep belief networks for hyperspectral image classification. IEEE Trans Geosci Remote Sens 55(6):3516–3530
9. Liu B, Yu P, Zhang X, Yu A, Fu Q, Wei X (2017) Supervised deep feature extraction for hyperspectral image classification. IEEE Trans Geosci Remote Sens 56(4):1909–1921
10. Qiu Q, Wu X, Liu Z, Tang B, Zhao Y, Wu X, Zhu H, Xin Y (2017) Survey of supervised classification techniques for hyperspectral images. Sens Rev 37(3):371–382
11. 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
12. Ranjan P, Kumar R, Girdhar A (2023) Recent CNN advancements for stratification of hyperspectral images. In: 2023 6th International conference on information systems and computer networks (ISCON), pp 1–5. IEEE
13. Zhan Y, Hu D, Wang Y, Yu X (2017) Semisupervised hyperspectral image classification based on generative adversarial networks. IEEE Geosci Remote Sens Lett 15(2):212–216

14. Ranjan P, Girdhar A (2024) Deep Siamese network with hand-crafted feature extraction for hyperspectral image classification. Multimed Tools Appl 83(1):2501–2526
15. Wang Y, Yao Q, Kwok JT, Ni LM (2020) Generalizing from a few examples: a survey on few-shot learning. ACM Comput Surv (CSUR) 53(3):1–34
16. Liu S, Shi Q, Zhang L (2020) Few-shot hyperspectral image classification with unknown classes using multitask deep learning. IEEE Trans Geosci Remote Sens 59(6):5085–5102
17. Mei S, Ji J, Geng Y, Zhang Z, Li X, Du Q (2019) Unsupervised spatial-spectral feature learning by 3d convolutional autoencoder for hyperspectral classification. IEEE Trans Geosci Remote Sens 57(9):6808–6820
18. Song W, Li S, Fang L, Lu T (2018) Hyperspectral image classification with deep feature fusion network. IEEE Trans Geosci Remote Sens 56(6):3173–3184
19. 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
20. Xu Y, Zhang L, Du B, Zhang F (2018) Spectral–spatial unified networks for hyperspectral image classification. IEEE Trans Geosci Remote Sens 56(10):5893–5909
21. Zhong Z, Li J, Luo Z, Chapman M (2017) Spectral–spatial residual network for hyperspectral image classification: a 3-d deep learning framework. IEEE Trans Geosci Remote Sens 56(2):847–858
22. Li Z, Ming L, Chen Y, Xu Y, Li W, Du Q (2022) Deep cross-domain few-shot learning for hyperspectral image classification. IEEE Trans Geosci Remote Sens 60:1–18
23. Xue Z, Zhou Y, Du P (2022) S3net: spectral–spatial Siamese network for few-shot hyperspectral image classification. IEEE Trans Geosci Remote Sens 60:1–19
24. Xue Z, Liu Z, Zhang M (2023) DSR-GCN: Differentiated-scale restricted graph convolutional network for few-shot hyperspectral image classification. IEEE Trans Geosci Remote Sens 61:1–18
25. Xue Z, Zhu T, Zhou Y, Zhang M (2022) Bag-of-features-driven spectral–spatial Siamese neural network for hyperspectral image classification. IEEE J Sel Top Appl Earth Obs Remote Sens 16:1085–1099
26. Huang L, Chen Y (2020) Dual-path Siamese CNN for hyperspectral image classification with limited training samples. IEEE Geosci Remote Sens Lett 18(3):518–522
27. Jia S, Jiang S, Lin Z, Xu M, Sun W, Huang Q, Zhu J, Jia X (2021) A semisupervised Siamese network for hyperspectral image classification. IEEE Trans Geosci Remote Sens 60:1–17
28. Jiang S, Jia S (2021) A 3d lightweight siamese network for hyperspectral image classification with limited samples. In: 2021 10th international conference on computing and pattern recognition, pp 142–148
29. Deng F, Pu S, Chen X, Shi Y, Yuan T, Pu S (2018) Hyperspectral image classification with capsule network using limited training samples. Sensors 18(9):3153
30. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504–507
31. BBengio Y, Lamblin P, Popovici D, Larochelle H (2006) Greedy layer-wise training of deep networks. In: Advances in neural information processing systems, vol 19
32. Chen Y, Lin Z, Zhao X, Wang G, Gu Y (2014) Deep learning-based classification of hyperspectral data. IEEE J Sel Top Appl Earth Obs Remote Sens 7(6):2094–2107
33. Wang X, Girshick R, Gupta A, He K (2018) Non-local neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7794–7803
34. Wang L, Wang L, Wang Q, Atkinson PM (2021) SSA-SiamNet: spectral–spatial-wise attention-based Siamese network for hyperspectral image change detection. IEEE Trans Geosci Remote Sens 60:1–18

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.

A Deep Learning Model to Generate Image Captions

Shashank Parmar, Raman Tyagi, Prince Kumar Dhankar

Department of Software Engineering, Delhi Technological University, New Delhi, India

DOI: https://doi.org/10.51244/IJRSI.2023.1011054

Received: 21 November 2023; Revised: 28 November 2023; Accepted: 04 December 2023; Published: 23 December 2023

ABSTRACT

How computers can automatically describe the substance of photographs using human language is a topic that interests us greatly. We choose to use the most advanced image caption generator currently available to obtain a deeper understanding of this computer vision topic. Show, attend and tell Visually attentive neural image caption generator [12]. Our machine learning-based neural network picture description generator is created in Python using the Pytorch ML framework. In our workflow, we've determined five key elements: Data preparation, a **convolutional neural network (CNN)** it helps in encoding, a **recurrent neural network (RNN)** it helps in decoding, a beam search to determine the best description, generation of sentences, and assessment make up the **R1-R6 framework**. The quality and correctness of the generated caption are evaluated using the BLEU-4 score. Each member of our group contributed equally to moving the project forward as we distributed the five elements mentioned above equally among ourselves. All five components have been completed successfully, and we can now use Kaggle Notebook to train our network. After the network has been trained and is performing satisfactorily, we continue to see the attention mechanism.

INTRODUCTION

The contemporary landscape of computer vision and machine learning is increasingly focused on endowing machines with the ability to autonomously generate textual descriptions for images. This multifaceted endeavor encompasses the recognition of intricate photographic scenes, involving the discernment of salient features and the transformation of visual depictions into coherent and simplified language. The potential of this undertaking manifests notably in the development of assistive technologies for individuals with visual impairments and the streamlining of automated captioning tasks in online environments.

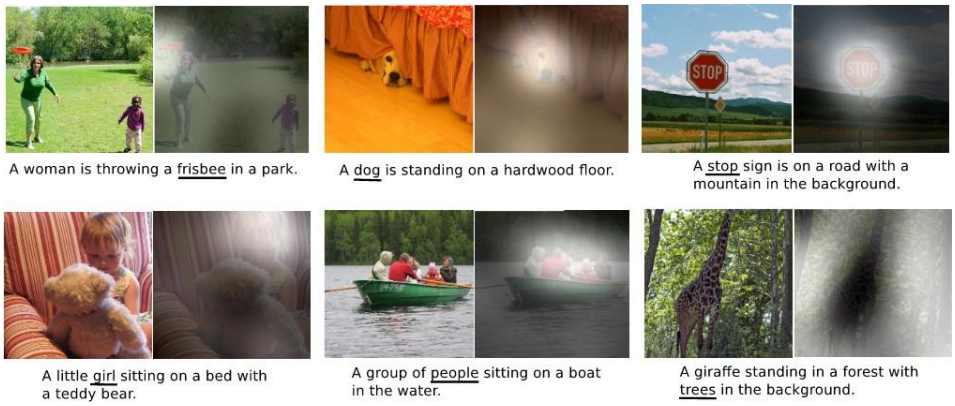


Figure 1. This figure is from Show, attend and tell visualization (adapted from [12])

While the pursuit of automatic image description has been a subject of considerable attention in recent decades, extant research publications have grappled with various challenges, including grammatical inaccuracies, cognitive incongruities, and the inclusion of extraneous content [5]. However, propelled by remarkable advancements in neural networks, recent endeavors have turned towards the integration of convolutional and recurrent neural networks to address these issues, yielding promising outcomes [2].

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume X Issue XI November 2023

Noteworthy contributions in this domain advocate the utilization of a deep convolutional neural network (CNN) in tandem with a recurrent neural network (RNN) to fulfill the objectives of image description. Building upon this foundation, a subsequent publication enhances the methodology by introducing an attention mechanism. Illustrated in Figure 1, the incorporation of a learnable attention layer empowers the network to selectively focus on specific regions of an image corresponding to each generated sentence. This augmentation represents a pivotal advancement, elevating the precision and contextual relevance of the generated textual descriptions. This discourse aims to synthesize and contribute to the ongoing dialogue within the scholarly community, offering insights into the evolution of methodologies for automated image description in the era of sophisticated neural network architectures.

METHODOLOGY AND ARCHITECTURE

To convert unformatted raw input data (including images and captions), we have created data preparation scripts; using a higher dimensional vector space to extract and encode picture characteristics using a pre-trained CNN architecture; a decoder that uses an LSTM-based recurrent neural network to translate encoded information into descriptions in natural language; Beam Search is used to determine the caption with the highest likelihood. For improving overall accomplishment, RNN uses an attention mechanism that allows it to recognise features from a part of the input image that has been deliberately emphasized. Below, a detailed explanation of each element of our generator pipeline will be provided.

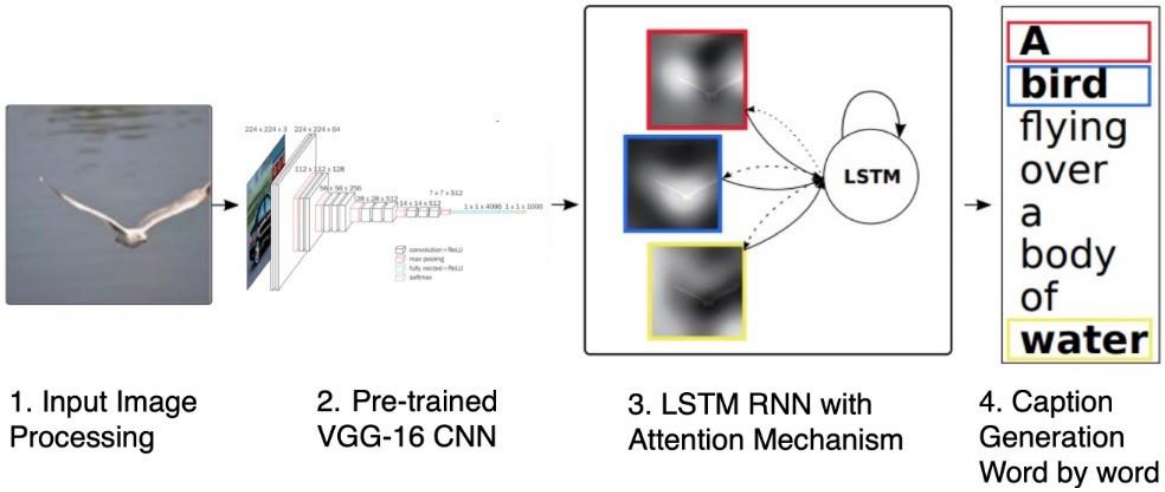


Fig 2. Picture depicts the photo's description generation process.

Data Sources

The COCO dataset [6], Flickr30k [8], and Flickr8k [3] are the three most famous picture caption training datasets in the Computer Vision research field. These datasets, which each contain 123,000, 31,000 and 8091 captioned photos, each have 5 descriptions for each image. Due to our limited storage and processing capability, we currently choose the Flickr8k dataset, which has the fewest photographs, as our main data source over the other two.

Data Preprocessing

Captions and images form the input data, thus should be preprocessed to place them in the correct formats for encoder and decoder, respectively. We must convert photos into the appropriate format because our image caption generating pipeline relies on a modern CNN network that has been pre-trained

PyTorch documentation states that pre-trained vgg-16 models assume input images to be normalized to fall between [0, 1] and take the form of 3-channel RGB photos with the shape (3, H, W), where H and W are

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume X Issue XI November 2023

assumed to be at least 224. Accordingly, data preprocessing comprises importing and scaling image data into (N, 3, 256, 256) dimensions as well as normalizing pixel values to be between [0, 1] with a mean of [0.485, 0.456, 0.406] and a standard deviation of [0.229, 0.224, 0.225].

Convolutional Neural Network

In order to later feed these visual qualities to RNNs, Prior to encoding, the CNN should first draw out graphical properties of various dimensions. VGG-16 & Res Net are frequently suggested image encoding algorithms. The PyTorch toolkit's pre-trained vgg-16 model is selected.

CNN is used in this project to encode the features rather than classifying the photographs. As a result, the network's end deleted the max pool and entirely connected layers. The dimensions of the supplied photo matrix are (N, 3, 256, 256), while the outcome of this new structure has (N, 14, 14, 512) dimensions. We also attached an adaptive 2D layer to the encoder's architecture to accommodate input photos of various sizes.

To cut down on computing costs, we removed gradient in our picture captioning architecture. We might get a better overall performance with some fine-tuning.

Soft Attention Mechanism

The next word in the description is produced by the network by following instructions from the attention mechanism to concentrate on a particular region of the image. The CNN outcome and the older state, which is upgraded after each reiteration, were combined to choose the attention region.

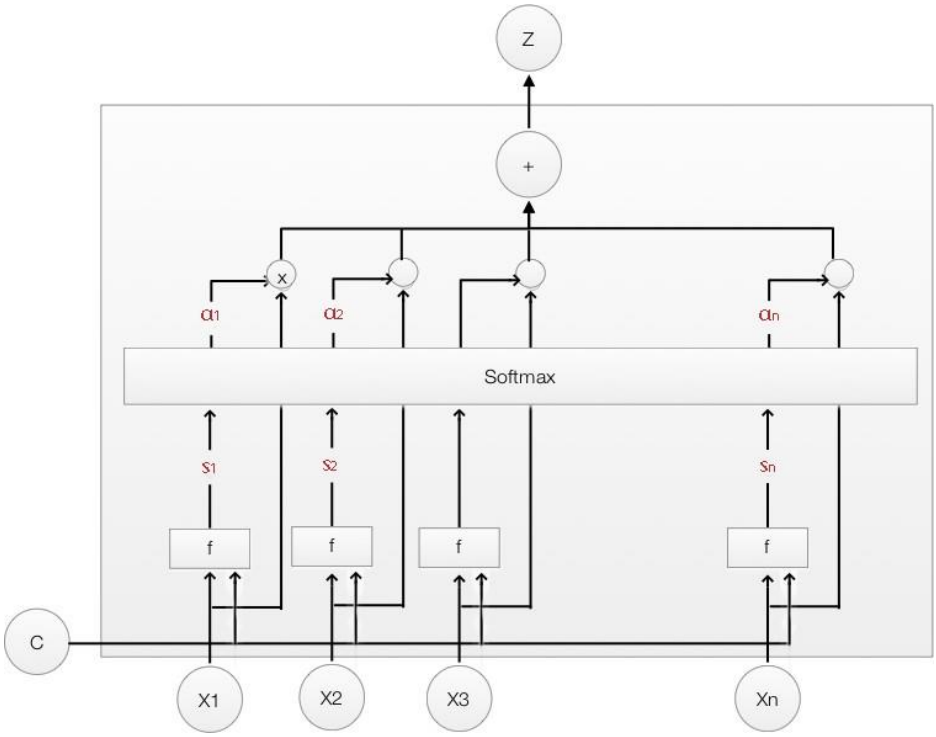


Figure 3. A demonstration of soft attention model [1]

When creating the "attention layer", we linearly transformed the encoder output as well as the historical state output. Total of the linearly activated outputs is the activation function, which is ReLU. The attention area (applying alpha to encoder output) and weight (the value of alpha) are returned.

The attention mechanism can be trained by back-propagation, in contrast to CNN without fine-tuning.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume X Issue XI November 2023

Later, a decoder using RNN uses the returned attention area.

Recurrent Neural Network (Decoder)

Using an RNN-LSTM that can produce words in a sequential manner, the decoder must construct word-by-word photo descriptions. The CNN photo features vectors that were formed during the data preprocessing stage and the encoded picture captions serve as the input for the decoder.

The RNN consists of an attention module we drafted and built, an LSTM cell module, & an LSTM cell module, in addition to four completely linked layers given by the PyTorch library for initializing the states of the LSTM cell and term dictionary.

As we get encoded photos and captions, we initially sort them according to the image's encoded key length. We only want to analyze the encoded photographs whose caption lengths are at least the same as the number of iterations in order to improve productivity and decrease time of training.

In order to create the attention-masked photos with a particular area highlighted, we first feed the encoded photos and the historical state of the LSTM network into the "Attention module" in each iteration of the LSTM network. Next state of the LSTM is obtained by concatenating the attention-grabbing visuals with all of the previous words' embedded captions. Fully linked layers can then forecast the likelihood of the present word embedding based on the present state & add to the word embedding prediction matrix.

Beam Search

The Beam Search in the last step of the Show and Tell paper, identifies the statement that has the greatest chance of recurrence given the input photo. The algorithm repeatedly considers the set of the top k sentences up to time a as candidates to generate sentences of size a + 1, being an informed search method, and keeps only the top k of them because this more closely resembles the chance of obtaining the highest cumulative probability score as cited in the paper. We applied beam search sizes with values 1, 2, 3, 4, 5, & the evaluation of the *Bilingual Evaluation Understudy* score appears to be more favorable for beam sizes 3 and 4. Therefore, we use 3 as size for beam search.

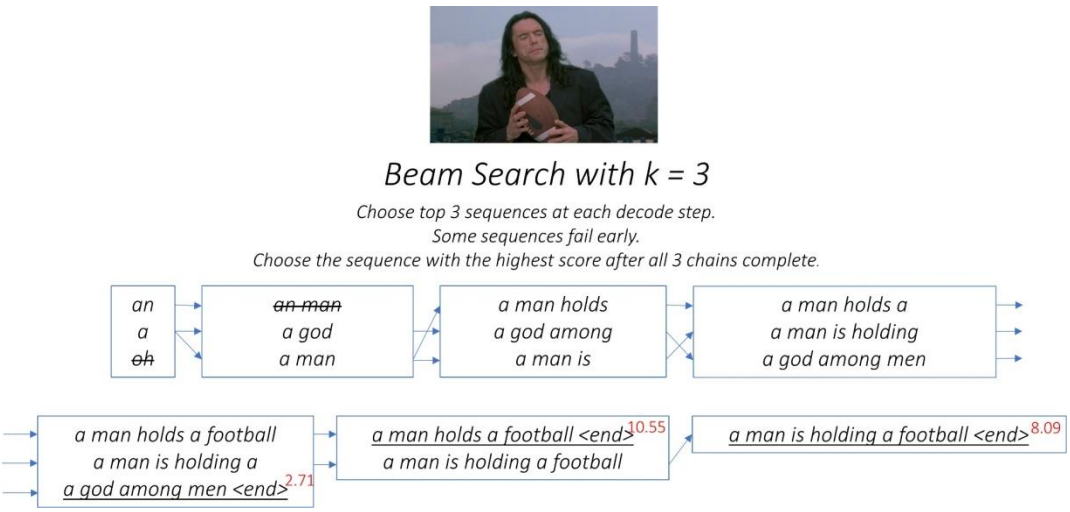


Fig 4. An example of Beam Search in our sentence generation.

From a github project, Figure 4 provides a useful illustration graphic [10]. This diagram exemplifies how beam search chooses the set of words with the global maximum rather than word with the highest probability at each stage (local maximum).

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume X Issue XI November 2023

Loss Function

As a series of probabilities of word occurrences make up the nature of our RNN output, we advise utilizing cross *entropy loss* to assess the *output's quality*. When a "classification model" generates a probability value between *0 and 1*, the most common and practical method to assess how well it is doing is to look at this probability value.

$$E_{entropy} = -\sum_{n}^{N} t_k^n \ln p_k^n \qquad (1)$$

EXPERIMENTS

In the course of our experimental investigations, numerous hyperparameters underwent systematic manipulation to refine the performance of our model. Initially, a learning rate of 4 × 10^-4 was applied during the training phase, resulting in significant oscillations in the training loss. To mitigate this issue, we implemented an adaptive learning rate schedule, denoted as the "lr scheduler," which incrementally elevated the learning rate by a factor of 0.99 after each epoch. This strategic employment of the scheduler effectively alleviated the pronounced oscillations, contributing to enhanced training stability.

Furthermore, the adjustment of the batch size constituted a pivotal aspect of our experimentation. Commencing with a batch size of 5 proved suboptimal, given the extensive composition of the dataset—specifically, the presence of hundreds of images within the "flicker 8k" dataset. This diminutive batch size hindered convergence of the loss function. Consequently, a progressive increase to a batch size of 50 was implemented, yet this expansion surpassed the 12-gigabyte (GB) memory constraint imposed by Kaggle's graphics processing unit (GPU). Ultimately, a judicious compromise was reached, and a batch size of 32 was selected, remaining within the GPU memory limitations while concurrently ensuring sufficient convergence of the loss function. This iterative optimization process underscored the necessity for a balanced configuration of hyperparameters to achieve both computational efficiency and optimal model performance.

EVALUATION

Although the pipeline has been established, which is not apparent whether it is entirely accurate or whether it can be taught in the manner for which it was designed. Our model must then be properly trained in order for it to learn. Another significant portion of this project that is yet unfinished is (R6) if the model has proven to be accurate. Sentence generation and performance assessment are the greatest ways to show performance of our description generator.

Evaluation Metrics

The *BLEU-1, BLEU-2, BLEU-3, and BLEU-4* metrics are the most often used to assess the level of accuracy of text creation in natural language processing tasks. Here, our main evaluation parameter was the 4-gram *BLEU* score *(BLEU-4)*. The following stage is to look into additional metrics (such METEOR) described in pertinent studies [11, 12] and attempt to make improvements to our model to make it more accurate.

Model Performance

Today, we have a 9.8 BLEU-4 score within 35 epochs (BLEU-4 is within range 0 to 100). Our model performs less well than state-of-the-art caption generators, as seen by the 19.5 BLEU-4 score reported in *Show, Attend, and Tell* on the Flickr 8k. We made the decision not to spend too much time and effort simply increasing our BLEU-4 score to match the findings in article,

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume X Issue XI November 2023

RESULTS

Data is translated into numbers using a lexicon of words, as was mentioned before. When data has ingested, the pipeline will also produce output in the encrypted format that must first be converted into English for human comprehension. The outcome of RNN is a series of word likelihoods, which is another crucial aspect. Selecting the word with the most likelihood at each decode stage in recurrent neural networks frequently yields less-than-ideal results. Beam Search, a popular method for determining the best pathways for decoding words in natural language, is used. which was covered in Section 2.6. The subtitles for a few test images can be found below.



Fig 5. An illustration of a caption that was created incorrectly.

Evidently, several of the automatically produced captions on the network omit significant image details, while others misidentify specific visual components. For instance, “boy in blue shirt is holding red pole in front of playground” can be seen in Figure 5’s top left image. The shirt in the image is identified by the model but the model fails to identify the correct color.

We were fascinated by how regularly and skillfully descriptions for pictures of dogs catching Frisbees were generated in test photographs. This is most likely because there are so many pictures showing dogs in training jumping to catch Frisbee. Furthermore, it’s probable that the pre-trained CNN is very familiar with how a dog looks.

Generated Captions

As seen in Figures 6, 7, 8, and 9, we were successful in producing captions that largely followed proper grammar rules and were legible by humans. The majority of the images give an accurate description of the objects in the scene, count how many times they appear, and give a logically sound verb to logically finish the sentence. Also effectively recorded are the object’s colors and the spatial interactions between items. Then, CNN is able to record the characteristics of items. Finally, conjunctions will be used by RNN and Attention Mechanism to logically join these words to form a whole phrase.



Fig 6. Dog jumping through an obstacle course

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume X Issue XI November 2023



Figure 7. Brown dog is running through the water

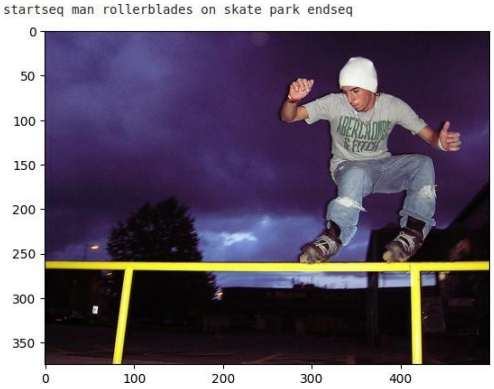


Fig 8. Man roller blades on skate park



Fig 9. Child in swimming pool

At the conclusion of this report, we have additionally added additional results (Figures 10 and 11).

Attention Mechanism Visualization

The capacity to visualize attention mechanisms enables us to comprehend where the area of the image is being focused when creating a certain phrase, which is crucial for enhancing the network's potential to grasp scenes. Each phrase in our visualization was made by layering a see-through mask over the original photo at each RNN node. As a result, the dark zone had little effect on the phrase that was formed, and the light region could be used to indicate the focal area.

CONCLUSION AND FUTURE WORK

Because mechanized picture descriptioning is still in its babyhood, there are various research projects now underway focusing on extra precise photo characteristic extraction and semantically superior phrase synthesis. We were only able to complete the tasks listed in the project proposal due to limited

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume X Issue XI November 2023

computational resources, and we had to use a smaller dataset (Flickr 8k). There might be enhancements if additional time is given. Because we used a pre-trained CNN network straight into our pipeline without making any further adjustments, the network did not respond to this specific training dataset. In order to enable fine-tuning and experiment with other CNN pre-trained networks, we anticipate getting a somewhat higher BLEU-4 score. Training on a mix of Flickr_8k, MSCOCO and Flickr_30k could also result in improvements. In general, the output will be more accurate the more divergent the training dataset. We can all agree that this study has piqued our curiosity in the potential applications of machine learning to computer vision, and we anticipate exploring these further in the future.

ACKNOWLEDGEMENT

Since we started the project from scratch, we ran into many execution challenges, many of which had to do with *PyTorch* syntax and usage conventions. With assistance from the *PyTorch developer community* and by consulting some pertinent open *Github* sources, we were able to resolve these difficulties.



Fig 10. Our caption generator produced image and caption pairs.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume X Issue XI November 2023



Fig 11. By using our image description generator, we can produce photos and captions both.

REFERENCES

1. Jonathan Hui Blog. https://jhui.github.io/2017/03/15/Soft-and-hard-attention.
2. Alex Graves. Generating sequences with recurrent neural networks. CoRR, abs/1308.0850, 2013.
3. Micah Hodosh, Peter Young, and Julia Framing image description as a ranking task: Data, models and evaluation metrics. J. Artif. Int. Res., 47(1):853–899, May2013.
4. Andrej Karpathy and Li Fei-Fei. Deep visual-semantic alignments for generating image IEEE Trans. Pattern Anal. Mach. Intell., 39(4):664–676, Apr. 2017.
5. Polina Kuznetsova, Vicente Ordonez, Alexander Berg, Tamara L. Berg, and Yejin Choi. Collective generation of natural image descriptions. pages 359–368, 2012.
6. Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Doll´ar, and C. Lawrence Microsoft coco: Common objects in context. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, Computer Vision – ECCV 2014, pages 740–755, Cham, 2014. Springer International Publishing
7. S. Liu and W. Deng. Very deep convolutional neural network based image classification using small training sample size. pages 730–734, Nov 2015.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume X Issue XI November 2023

8. Bryan A. Plummer, Liwei Wang, Chris M. Cervantes, Juan C. Caicedo, Julia Hockenmaier, and Svetlana Lazebnik. Flickr30k entities: Collecting region-to-phrase correspondences for richer image-to-sentence Int. J. Comput. Vision, 123(1):74–93, May 2017.
9. PyTorch. https://ml-cheatsheet.readthedocs.io/en/latest/loss_functions.html.
10. Sagar https://github.com/sgrvinod/a-PyTorch-Tutorial-to-Image-Captioning
11. Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Show and tell: A neural image caption generator. CoRR, abs/1411.4555, 2014.
12. Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron C. Courville, Ruslan Salakhutdinov, Richard S.Zemel, and Yoshua Show, attend and tell: Neural image caption generation with visual attention. CoRR, abs/1502.03044, 2015.

The Visual Computer
https://doi.org/10.1007/s00371-023-03256-4

ORIGINAL ARTICLE

A human activity recognition framework in videos using segmented human subject focus

Shaurya Gupta[1] · **Dinesh Kumar Vishwakarma**[2] · **Nitin Kumar Puri**[1]

Accepted: 24 December 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
Automating tasks through human activity recognition in video data has become increasingly vital. Deep learning has yielded versatile activity recognition systems applicable in surveillance, healthcare analysis, sports, and human–computer interaction. Despite various proposed video-based activity recognition techniques over the years, the reliance over RGB frames, accompanied by other modalities like joint locations and depth maps, often proves less effective compared to multimodal methods. In response to this challenge, our paper introduces a competitive approach for identifying human activity in video frames. Leveraging a Convolutional Long Short-Term Memory (Conv-LSTM) network and a novel pre-processing step involving a Human Segmentation network, our method accentuates human subjects in each frame using segmentation maps. These highlighted frames undergo further processing through Convolutional Neural Networks (CNNs) to learn feature vectors, without including other modalities with RGB frames directly. The learned features are then subjected to Long Short-Term Memory (LSTM) units for comprehending sequential video data and drawing meaningful inferences. The proposed methodology undergoes rigorous testing on three publicly available datasets—KARD, MSR Daily Activity, and SBU-Interactions. Remarkably, our approach outperforms similar state-of-the-art methods, achieving benchmark accuracy scores exceeding 98% on MSR Daily Activity and 99% on KARD and SBU-Interactions datasets. In essence, our method not only provides a competitive solution for human activity recognition in video frames but also contributes to advancing the field by integrating Conv-LSTM networks and innovative pre-processing techniques. The comprehensive evaluation on multiple datasets underlines the robustness and superior performance of our proposed approach.

Keywords Convolutional network · Deep learning · Human activity recognition · Image segmentation · Bi-directional LSTMs

1 Introduction

Human Activity Recognition (HAR) is the task of discerning human activities from sequential input data, achieved through various means such as video frames capturing human actions [1] or sensors embedded in smartphones and smartwatches [2]. In this paper, we focus on image processing for activity recognition, proposing a novel vision-based system. The applications of activity recognition are widespread, encompassing sports action analysis, anomalous behavior recognition, intelligent surveillance systems, and medical patient monitoring, etc [3].

Despite its potential, video based HAR is fraught with challenges arising from environmental complexities, occlusions, and the similarity between different actions. However, recent advancements in deep learning techniques and visual data have fueled substantial growth in research and application in this domain. Alongside traditional RGB video frames, novel action recognition techniques have incorporated additional features/modalities such as human joint locations [4], frame depth [5], and silhouette maps [6]. Figure 1 illustrates some of these modalities. While these techniques have enhanced baseline accuracy, their reliance on additional features often demands expensive equipment like Microsoft

✉ Dinesh Kumar Vishwakarma
dinesh@dtu.ac.in

Shaurya Gupta
shaurya62000@gmail.com

Nitin Kumar Puri
nitinkumarpuri@dtu.ac.in

[1] Department of Applied Physics, Delhi Technological University, Delhi, India

[2] Department of Information Technology, Delhi Technological University, Delhi, India

Published online: 06 February 2024

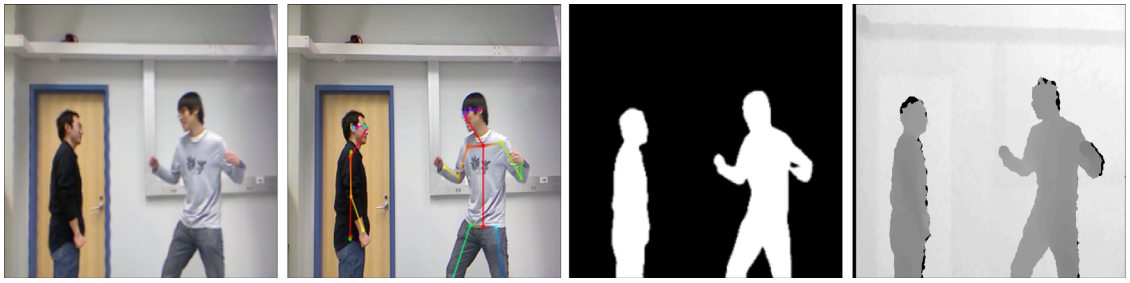


Fig. 1 Different modalities used as input for Human Activity Recognition task. (Sample frame from one of the datasets)

Kinect for depth estimation and joint coordinate determination through specific estimation algorithms [7].

Our work acknowledges the necessity of capturing additional image information to enhance activity recognition. We introduce human segmentation masks to highlight the human subject in each video frame, forming a competitive learning pipeline. Image segmentation, which classifies pixels in an image into specific categories, aids in identifying regions where the subject is highlighted against the background. Given the sequential nature of image data, we employ Convolutional Neural Networks (CNNs) to extract features and Long Short-Term Memory (LSTM) like architectures to understand the relation between video frames. Through this, we do predictive modeling over highlighted humans to identify the concurrent activity happening in the video. Further, this helps the network to process the input as a single RGB modal input rather than concatenation or addition of separate ones, thus also reducing the computational overhead to some extent.

In contrast to methods employing different network streams to process additional input features alongside RGB frames [8], our approach generates a human-focused highlighted video frame through segmentation-based preprocessing, enhancing learning for the performed activity. The experiments yield substantial results compared to the use of plain RGB frames through the proposed deep learning architecture, rivaling techniques utilizing other multi-modal input features like joint location or depth maps. The accuracy and self-reliance of the framework over segmentation on RGB frames make this work applicable in real-world scenarios.

The significant contributions of this paper are as follows:

- A novel pre-processing pipeline is proposed for HAR, leveraging human segmentation. This innovative approach enhances the focus on humans within video frames, improving the interpretability of activity patterns.
- A training pipeline compromising of a Conv-LSTM like architecture using bi-directional LSTMs to capture both forward and backward temporal dependencies, enhancing its capacity to understand the temporal dynamics of human activities using segmentation pre-processed inputs.
- The proposed architecture achieves outstanding results on three publicly available datasets, namely KARD, MSR Daily Activity, and SBU-Interactions. Comparative evaluations against state-of-the-art techniques underscore the efficacy and robustness of the proposed approach.

The rest of the paper is organized as follows. Section 2 discusses the related works in human activity recognition; Sect. 3 explains the proposed methodology with a brief explanation for segmentation and pre-processing technique for Conv-LSTM input. Finally, Sect. 4 shows the experiments and results comprising of details on datasets and training details with a comparison of other state-of-the-art methods at the end.

2 Related works

Artificial Intelligence (AI) has gained significant attraction in mainstream applications within the last few years. With the advent of parallel computing through GPUs, AI, particularly deep learning, has seen many applications like violence detection [9], sentiment analysis [10], crowd anomaly detection [11], Gait Recognition [12], and many more. Activity recognition has mainly 2 instances either through video or wearable sensor data [13, 14]. The considered task in this paper, i.e., Human Activity Recognition in video frames, is also worked upon using current deep learning techniques. Before deep learning, most methods relied on the hand-crafted features understood from a video frame before proposing any algorithm to classify the task as a specific activity. Current works in video activity recognition generally rely on using more features like depth images for each video frame [15], skeleton information/ joint locations [16], or forming some spatial information vector through the

combination of RGB video frames. Alternatively, some techniques employ 3D-CNN architectures to convolve over the entire temporal sequence, although such methods demand substantial processing power and are susceptible to overfitting [17].

Most of the works tried to implement the CNN-based Activity Recognition system but still didn't receive any extraordinary accuracy [18]. The primary reason for this was data-dependency, i.e., quality of data with human activity. Thus, many techniques tried to improve their accuracy by using additional features to improve the features understanding through the video data. For instance, Yan et al. [19] proposed using a physical and geometric embedding network for recognition through skeleton data and edge information to form a two-stream network and combine their results for activity recognition. Other than these sequential features, many researchers develop new spatial information through a combination of data. Simonyan et al. [8] gave a two-stream network that, apart from RGB, considered the video frames' optical flow to finally fuse the results for activity prediction. Singh et al. [20] generated a dynamic motion image from the set of video sequences for an activity sent to a convnet to detect human activity through a two-stream network taking RGB frames and dynamic-motion images, respectively, as input. Their work also includes a CNN-LSTM based network to predict the expected activity through the video frames. The results from these two networks were fused to get a higher accuracy score. On the other hand, Dhiman et al. [5] applied a similar fusion technique, with the RGB data replaced by the depth estimated maps which are cropped to region-of-interests. The results of this depth map processed convnet are combined with a dynamic motion image processed network to provide some state-of-the-art results. Zebhi et al. [21] also form a spatial feature using a gait history template that works on accumulation of region of motion into a single image; these features are then, like most other methods, processed through CNNs for activity prediction. Kapoor et al. [6] use silhouettes features to extract the occurring activity. Still, the technique is dependent on the given data distribution where the background complexity is very minimal, making it easy to extract human segments.

Through a comprehensive review, it is evident that the inclusion of additional image features, alongside RGB input, significantly enhances the accuracy of the activity recognition pipeline. While human silhouettes have been employed in prior works, they were either integrated into two-stream networks or used to construct spatial images. This paper introduces a novel technique that not only generates and applies silhouettes to focus on human subjects in complex environments but also excludes occlusions, making humans the primary target for learning through successive frames. The proposed highlighting technique presents a distinctive approach to enhance input video frames, contributing to the ongoing advancements in Human Activity Recognition.

3 Proposed architecture

While predicting activity on any video, the key observations are made on human movement and their interaction with the environment captured through the consecutive frames. A majority of the techniques can be observed to either form some spatial information vector after processing through encoding techniques (generally convnets) from the set of given video frames or extract some human characteristics like joint position maps, silhouettes, etc. These considered features are then either sent through a two-stream network by inputting original video frames along with the new features to predict the possible ongoing activity. Considering the importance of human subject(s) present in each video frame throughout the temporal stream, we introduce a segmentation aided pre-processing step to experiment and show the improvements of providing human subject area focused video frames as inputs to a deep learning architecture and detect the concurrent activity. The proposed technique relies on a deep learning-based segmentation technique to detect the human subject mask(s) in each passing frame with some pre-processing before sending data to a Convolutional-LSTM (Conv-LSTM) based architecture for activity prediction.

Hence, through this work, a human segmentation aided pipeline is proposed, training a separate segmentation network to predict segment masks for the human subject in each video frame with some further pre-processing to reduce the foreground information from the image, thus making the human subject appear in much higher focus.

Once the human subjects appear in higher focus than the relative background, the video sequence frames are fed through an image-net [22] pre-trained Convolutional Neural network (CNN) to encode the video sequence frames into feature vectors, reducing the spatial dimensions drastically, to be easily understood by a recurrent neural network.

The convolution processed feature vectors are then fed to Bidirectional-LSTMs to understand the relation between the video frames considering both forward and backward information at a current frame-i. This bi-directional action helps in understanding the relation not only from previous video frames but also from frames in the future. Even by evidence from research of recent works, it is inferred that Bidirectional LSTMs provide much better results than standard LSTMs, considering only past information at any point [23].

Finally, the Bi-directional LSTMs features act as inputs for a fully connected neural network, which maps the network nodes to human activity, classified based on the softmax

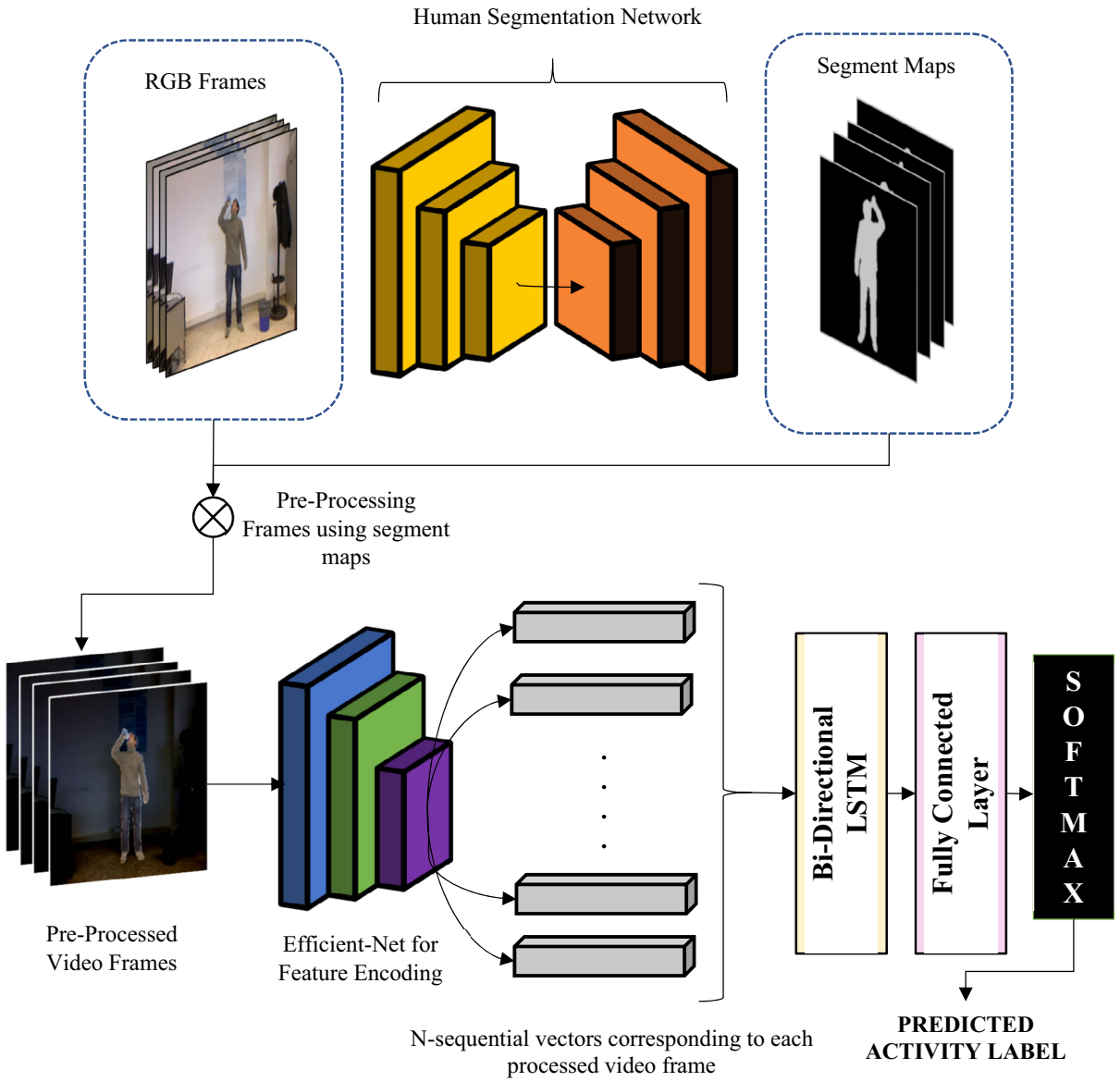


Fig. 2 Proposed segmentation guided architecture for human activity recognition

function used in the last layer. The proposed framework can be observed in Fig. 2.

3.1 Segmentation technique

Segmentation refers to classifying each pixel in a given image as a particular class. This section explains the employed segmentation architecture to generate the human masks for video pre-processing. Figure 3 shows the encoder-decoder architecture for image segmentation achieved using a DenseNet [24] encoder and a custom decoder aided by intermediate concatenations connection between encoder-decoder networks. This architecture is very similar to U-net [25] but relies more on encoder learning since information is passed on to the decoder via concatenation connections. The given segmentation architecture is trained on Supervisely Person's dataset [26] for its wide variety of human region masks. The considered architecture provides excellent results on unseen images, as observed from Fig. 4. Each of the video frame considered is read at a dimension of 256 × 256, Thus the segmentation architecture is trained on same dimension. For quick convergence, the DenseNet is also considered to be pre-trained on the image-net [22] dataset. The network consisted of 338 layers and 7Million parameters, comprising both encoder and decoder layers. The model's training is done for a mere 30 epochs at a learning rate of 1e-3 using Adam optimizer. Further, it is evident from Fig. 3 that the results are sufficient for the considered pipeline and video

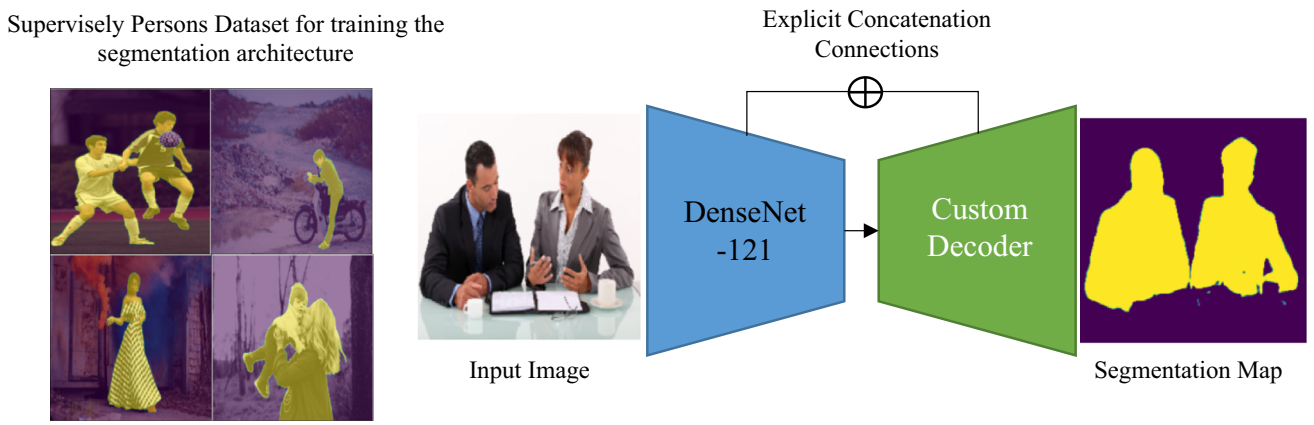


Fig. 3 Segmentation network for producing human masks in input video frames. The yellow region represents the human segmentation area. The structure is similar to U-net varying in its encoder structure



Fig. 4 Results produced by the implemented segmentation network over the considered activity datasets and a randomly selected image

pre-processing. An important observation here is that when humans have small objects like a glass, can, or phone, these regions are also included in the segmentation map since they are close to the subject's hand and available in the dataset. To improve activity recognition of humans, we only consider human segmentation and not for some particular objects that humans might use to interact with. Further, human-to-human interaction is perfectly captured in one of the datasets, as observed in Fig. 4.

3.2 Dataset pre-processing

Applying the image segmentation architecture observed in Fig. 3, we obtain the human masks (or silhouettes), which are used to improve the human presence throughout the image. This section teaches the pre-processing technique applied to the input video frames through segmentation masks, making the human subject appear as a prime focus. To reduce the foreground information, we decrease the pixel intensity in the regions where the mask has pixel values close to 0, i.e., the segmentation map classifies the region as non-human. To reduce the pixel intensity, we use the mean of each color channel separately, i.e., RGB. Based on this mapping, we form an image where the human subject is brighter than the relative background. The operation of reducing the pixel intensity can be inferred from Eq. (1) below.

$$\text{Image}_{\text{pre - process}} = \sum_{k}^{\text{RGB}} \sum_{i}^{256\times256} \begin{cases} \max\left(I\left[i,k\right]-\mu\left(k\right),0\right), if\, M\left[i\right]=0 \\ I\left[i,k\right], if\, M\left[i\right]=1 \end{cases} \tag{1}$$

where I represent the original image, M is the mask predicted by the segmentation network, and μ is the mean pixel intensity for each image channel Red, Green, and Blue. With the given pre-processing, the image gets divided into low pixel densities and higher ones making it easily distinguishable for the presence of a human subject in the video frames. Thus,

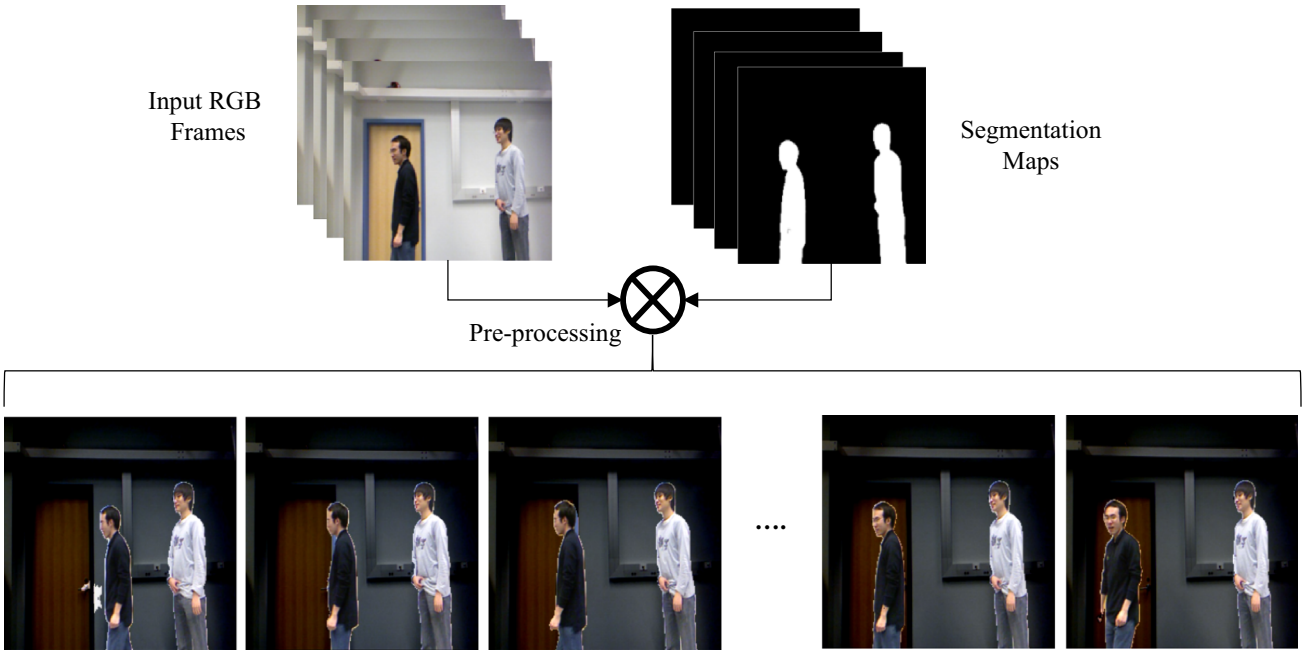


Fig. 5 RGB input frame pre-processing using segmentation maps produced by the encoder-decoder network. The processed frames reduce the foreground information and make human subject the key focus

we pre-process each video frame based on the above equation to get an enhanced human subjected image, as shown in Fig. 5. The results in Fig. 5 can be seen to consider the human subject quite efficiently for passing through the proposed Convolutional architecture and predicting the ongoing activity.

3.3 Proposed Conv-LSTM architecture

The architecture proposal following the pre-processing utilizes a pre-trained convolutional encoder as Efficient-Net [27] to obtain reduced spatial encodings to be decoded using a sequential network for inference on the activity throughout the video frames. To understand the sequential information, we use Bi-directional LSTMs for improved accuracy observed in previous techniques employed. The given work starts predicting activity after collecting 25 frames within a video sequence. Each considered frame is fixed at a dimension of 256×256. Once frames are grabbed from a video, they are passed through the segmentation network for improved human focus and sent through the network architecture previously observed in Fig. 2. Each frame gets its encoding and is passed onto the bi-directional LSTM. The bi-directional LSTM results are then provided to the Fully Connected layer for learning and finally mapping with a softmax layer to predict the possible output category over the available classes.

3.3.1 Efficient net

The selection of the encoding method is crucial for activity recognition. Since we are encoding the visual information for inferencing via a sequential neural network, the data needs to be optimally converted into feature vectors to decode the activity from the sequential data. Most of the convolutional neural networks (CNN) have a large scale, i.e., they use many parameters and rely on increasing the number of layers to improve their accuracy. To get better efficiency with minimal parameters, we use Efficient-Net, introduced by Google, which applies compound scaling to enhance accuracy while using a slight increase in parameters. A general idea of scaling can be observed from Fig. 6, where the Efficient net proposes to apply a combination of scaling techniques to increase the accuracy. Scaling differently about different dimensions results in better learning since scaling in each aspect affects the architecture differently.

Table 1 displays the parameter efficiency of a few variants of Efficient Net compared to other Benchmarked architecture. Further, Efficient-Net employs Mobile Inverted bottlenecks [32] with further addition of squeeze and excitation optimization [33]. Considering the results in Table 1, we apply Efficient-Net B2, pre-trained on Image Net for encoding the video frames. As mentioned before the network takes input a $25 \times 256 \times 256 \times 3$ input which is preprocessed with masks of same height and width produced by the segmentation network. The input is reduced to a vector of $25 \times 8 \times 8 \times 1408$ which gets further reduced to

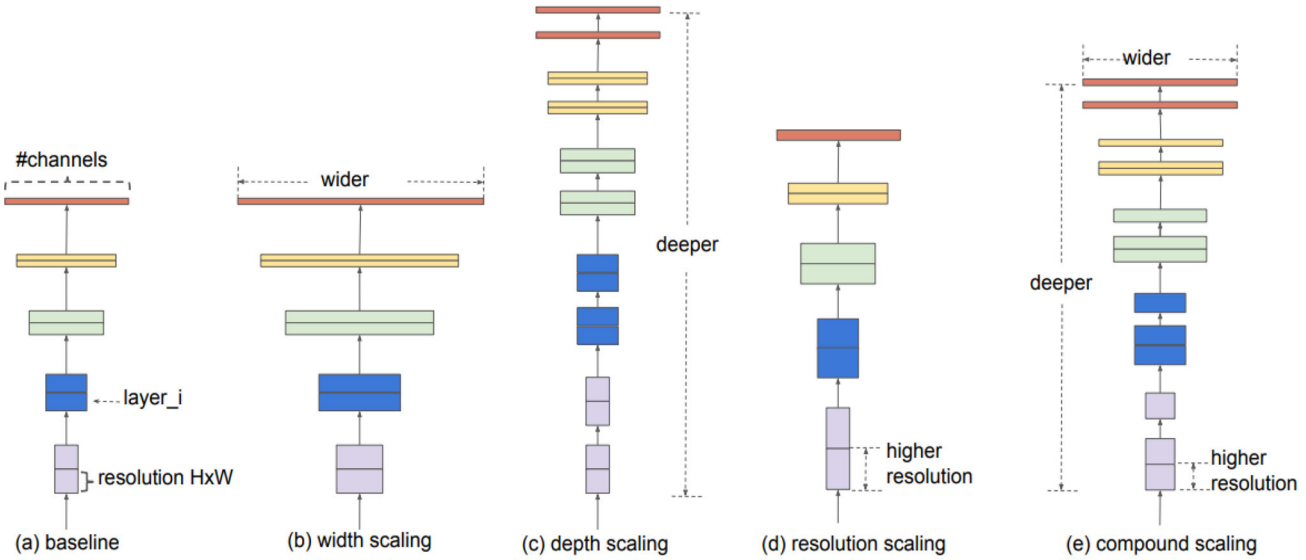


Fig. 6 Scaling Technique used by Efficient-Net. Here, **a** is the general architecture where **b-d** are the various scaling techniques applied in only one dimension. **e** Is the compound technique used by Efficient-Net to improve its accuracy

Table 1 ImageNet performance results between difference state-of-the-art models and EfficientNet

Model	Top-1 Acc (%)	Top-5 Acc (%)	#Params	#FLOPs
EfficientNet-B0	**77.1**	**93.3**	**5.3 M**	**0.39B**
ResNet-50 [28]	76	93	26 M	4.1B
EffieicentNet-B1	**79.1**	**94.4**	**7.8 M**	**0.7B**
ResNet-152 [28]	77.8	93.8	60 M	11B
DenseNet-264 [24]	77.9	93.9	34 M	6B
Inception-v3 [29]	78.8	94.4	24 M	5.7B
Xception [30]	79	94.5	23 M	8.4B
EfficientNet-B2	**80.1**	**94.9**	**9.2 M**	**1B**
Inception-resnet-v2 [31]	80.1	95.1	56 M	13B

Bold value indicates the accuracy of EfficientNet varients

25×1408 using global average pooling. These feature vectors are then fed to Bi-directional LSTM for understanding relation between given frame encodings.

3.3.2 Bi-directional LSTM

Generally, Recurrent Neural Networks (RNN) are used for processing sequential data and learning the underlying features for tasks like classification and regression. But RNNs face vanishing gradients problem during training when containing long sequential input data; LSTMs (Long Short-Term Memory) [34] recurrent neural networks solve this issue and process long inputs easily for desired task execution. A basic LSTM structure can be inferred from Fig. 7. They appear like a chain-like neural net consisting of different gates within each memory unit to process information via multiplication and addition of different gates, namely: input, forget, and output. For a hidden LSTM cell following operations are carried out as shown in Eqs. (2–7):

$$i_t = \sigma\left(W_i.\left[h_{t-1}, x_t\right] + b_i\right) \tag{2}$$

$$f_t = \sigma\left(W_f.\left[h_{t-1}, x_t\right] + b_f\right) \tag{3}$$

$$\widehat{C}_t = \tanh\left(W_c.\left[h_{t-1}, x_t\right] + b_c\right) \tag{4}$$

$$c_t = f_t * c_{t-1} + i_t * \widehat{C}_t \tag{5}$$

$$o_t = \sigma\left(W_o.\left[h_{t-1}, x_t\right] + b_0\right) \tag{6}$$

$$h_t = o_t * \tanh(c_t) \tag{7}$$

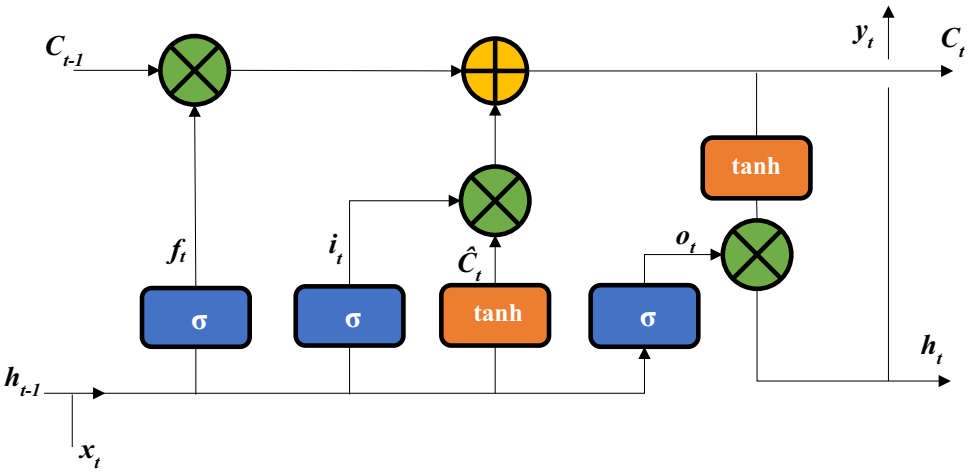


Fig. 7 Basic LSTM structure, where the symbols have their meaning as addition and multiplication respectively

where, i_t is the input gate, h_{t-1} is previous unit's output, σ is the activation function applied, W and b denotes the weight and biases of specific gates. $\hat{C}_t$ is referred to as input modulation gate for updating the cell state c_t.

While backpropagating through each of the LSTM units, the multiplication is carried out through various forget gates, which makes LSTMs less prone to vanishing gradients. Throughout these steps, an LSTM unit identifies which information is needed to be kept and which needs to removed and finally combines all the information to get that particular unit's outputs and cell state c_t.

Though LSTMs are better than standard RNNs, they only consider information in one direction i.e., only past information of features is considered. In human activity recognition the future information can be as important as the past one and help infer better results during training. Thus, we apply Bi-directional LSTMs [35] in our architecture to improve the classification results by giving current LSTM unit information from previous as well as next units. Figure 8 shows the Bi-directional LSTM for understanding the exchange of inputs and outputs. Since we consider both forward and backward LSTMs units in bi-directional, we get the following equations (Eqs. 8–10):

$$\overrightarrow{h_t} = \sigma\left(W_{\vec{h}}.\left[x_t, \vec{h}_{t+1}\right] + b_{\vec{h}}\right) \tag{8}$$

$$\overleftarrow{h_t} = \sigma\left(W_{\overleftarrow{h}}.\left[x_t, \overleftarrow{h}_{t+1}\right] + b_{\overleftarrow{h}}\right) \tag{9}$$

$$y_t = W_y.\left[\overrightarrow{h_t}, \overleftarrow{h_t}\right] + b_y \tag{10}$$

where, $\overrightarrow{h_t}$ is the forward hidden sequence of LSTM units and $\overleftarrow{h_t}$ is the backward hidden sequence and y_t is the output sequence of the bi-directional LSTM. Generally, σ i.e., the activation used is sigmoid throughout the equations used.

4 Experiments and results

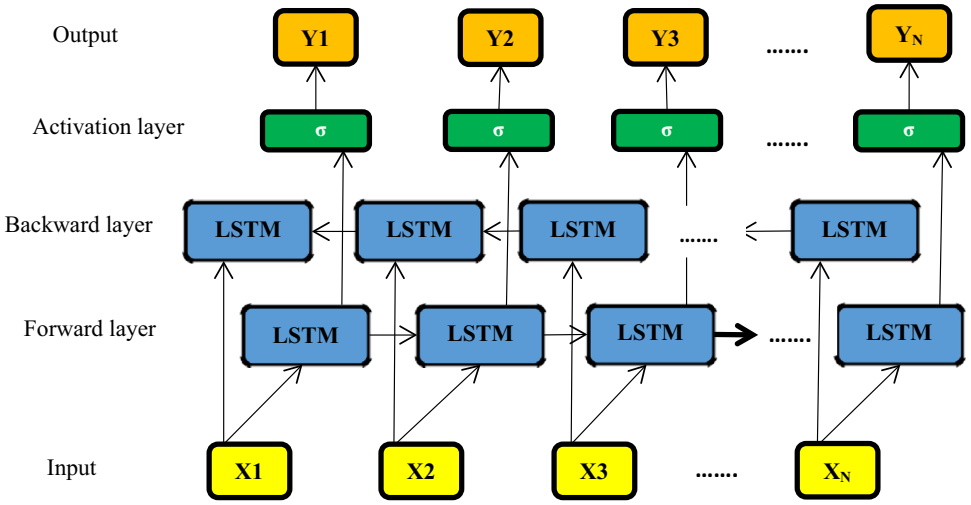
For training and implementation of the proposed technique Tensorflow framework was used in python. All the experimentations were carried out on a system consisting of NVIDIA 16GB Tesla P100 GPU with 3500 CUDA cores, CPUs @ 2.2Ghz, and 14GB RAM. This section discusses the datasets used for experimentation, training details, and result comparison with some recent benchmark approaches over each considered dataset.

4.1 Datasets

To test and analyze the performance of the proposed architecture, experiments are conducted over some benchmark datasets, namely, KARD [36], MSR Daily Activity [37], and SBU Interactions [38]. Each dataset provides either video or extracted frames. In case videos are present, the respective frames are extracted to a directory of video name. Further, these datasets contain information regarding depth maps and human body key-points which are not required by us since we are relying on the pre-processing step to enhance the human subject in each frame.

4.1.1 Kinect activity recognition dataset (KARD)

Gaglio et al. [36] introduced KARD consisting of RGB video frames along with their depth maps and corresponding joint positions. The dataset consists of total 18 activities namely: Horizontal arm wave (K1), High arm wave (K2), Two hand wave (K3), Catch Cap (K4), High Throw (K5), Draw X (K6), Draw Tick (K7), Toss Paper (K8), Forward Kick (K9), Side Kick (K10), Take Umbrella (K11), Bend (K12), Hand Clap (K13), Walk (K14), Phone Call (K15), Drink (K16), Sit Down (K17), and Stand Up (K18). Each activity is performed



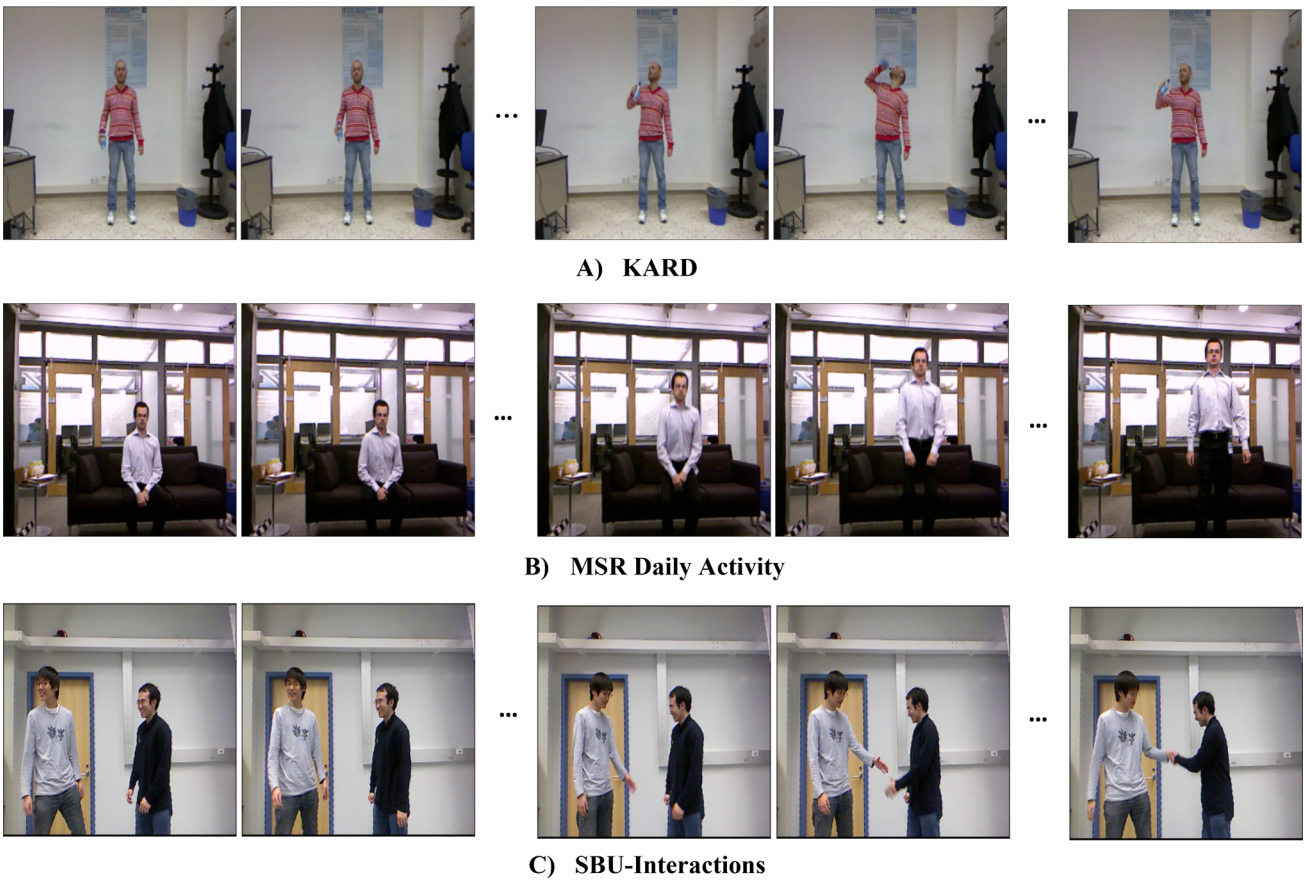


Fig. 8 The architecture of bidirectional LSTM

Fig. 9 Sample images from all three datasets (KARD (**a**), MSR daily activity (**b**), SBU-interactions (**c**)). The video frames for each dataset have the given environment as shown in these images

by 10 different human subjects with 3 times repetition. Sample video frames of drinking activity can be observed in Fig. 9a.

4.1.2 MSR daily activity dataset

Wang et al. [37] provided this data comprising of some common daily activities in a closed room environment, as visible in Fig. 9b showing Standing Up activity. This dataset additionally contains depth maps for each RGB video. Here, we have tendifferent human subjects carry out 16 daily activities twice, namely, Drink (M1), Eat (M2), Read Book (M3), Call Cellphone (M4), Write on paper (M5), Use Laptop (M6), Use Vacuum Cleaner (M7), Cheer Up (M8), Sit Still (M9), Toss Paper (M10), Play Game (M11), Lie Down on Sofa (M12), Walk (M13), Play Guitar (M14), Stand Up (M15), and Sit Down (M16).

4.1.3 SBU interactions dataset

Yun et al. [38] proposed this human interaction dataset containing different basic activities involving 2 humans subjects. This dataset also contains additional depth maps for each frame and corresponding 3D coordinates of human joints. In this dataset, we have 8 activities namely, Approaching (S1), Departing (S2), Kicking (S3), Punching (S4), Pushing (S5), Hugging (S6), Shaking Hands (S7) and Exchanging (S8). These activities are performed by 7 participants with randomly selected two participants for performing the activity. Sample video frames can be observed in Fig. 9c, showing Shaking Hands activity.

4.2 Training details

The proposed technique is end-to-end trained. The weights of the considered Efficient Net B2 are pre-trained on Image-Net, and further fine-tuned for each of the three considered datasets separately. The Backpropagation through time (BPTT) algorithm is applied to get the optimal LSTM weights, and since CNN architecture is fine-tuned, the backpropagation pass is used to find the optimal CNN weights. All the layers used in fully connected layers use ReLU activation (Eq. 13) with final classification using softmax (Eq. 12) to identify the output results. For optimization of the proposed architecture RMSProp optimizer is used at a learning rate of $1e-3$ (Eq. 14–16). Since we are dealing with a classification task, we apply categorical cross-entropy loss as signified in Eq. 11.

Loss Function: Categorical Cross Entropy

$$\text{Loss}(L) = -\frac{1}{m}\left(\sum_{i=1}^{m} y_i \log\left(\hat{y}_i\right)\right) \tag{11}$$

Softmax Function:

$$\text{softmax}(x_i) = \frac{\exp(x_i)}{\sum_j x_j} \tag{12}$$

ReLU activation:

$$\text{ReLU}(z) = \max(z,\, 0) \tag{13}$$

RMSProp Optimizer Algorithm:

$$g = \frac{1}{m}\nabla_\theta \sum_i L\left(f\left(x^{(i)};\, \theta\right),\, y^{(i)}\right) \tag{14}$$

$$\boldsymbol{r} = \rho \boldsymbol{r} + (1-\rho) g \odot g \tag{15}$$

$$\theta = \theta - \frac{\varepsilon}{\sqrt{\delta + \boldsymbol{r}}} \odot g \tag{16}$$

where, m is the batch size which is selected as 8 in our use case, θ is the weight parameter being updated, y is the ground truth classification variable, $\hat{y}$ is the predicted classification, x or z are the input to function given in equations, g is the computed gradient, r is the accumulated squared gradient over the batch, $\odot$ is the element-wise multiplication operator. A small value $\delta(1\text{e}-7)$ is added to ensure denominator doesn't reduce to zero. ε is the learning rate and ρ is the decay rate.

To avoid the occurrence of overfitting during training, a drop-out mechanism is adopted. The proposed architecture is trained for a mere 30 epochs using the adopted optimizer. The training over each dataset can be observed in Fig. 10. The quick convergence over the given loss function can be attributed to fine-tuning the pre-trained Conv-Net on the given activity dataset with the training at the end for each dataset showing minimal overfitting.

4.3 Results and discussion

Finally, the performance of the model on training and testing data is done using the following evaluation metrics namely: Confusion Matrix, Accuracy, Precision and Recall. A brief explanation on these metrics is summarized as follows:

1. **Confusion matrix:** It is a method to identify prediction quality of a classifier where we identify the following terms for categorical predictions:
2. **True Positives (TP):** For a particular category to be classified if the ground truth and the classifier prediction are both positive.
3. **True Negatives (TN):** For a particular category, if ground truth is not present and the classifier anticipates the same.

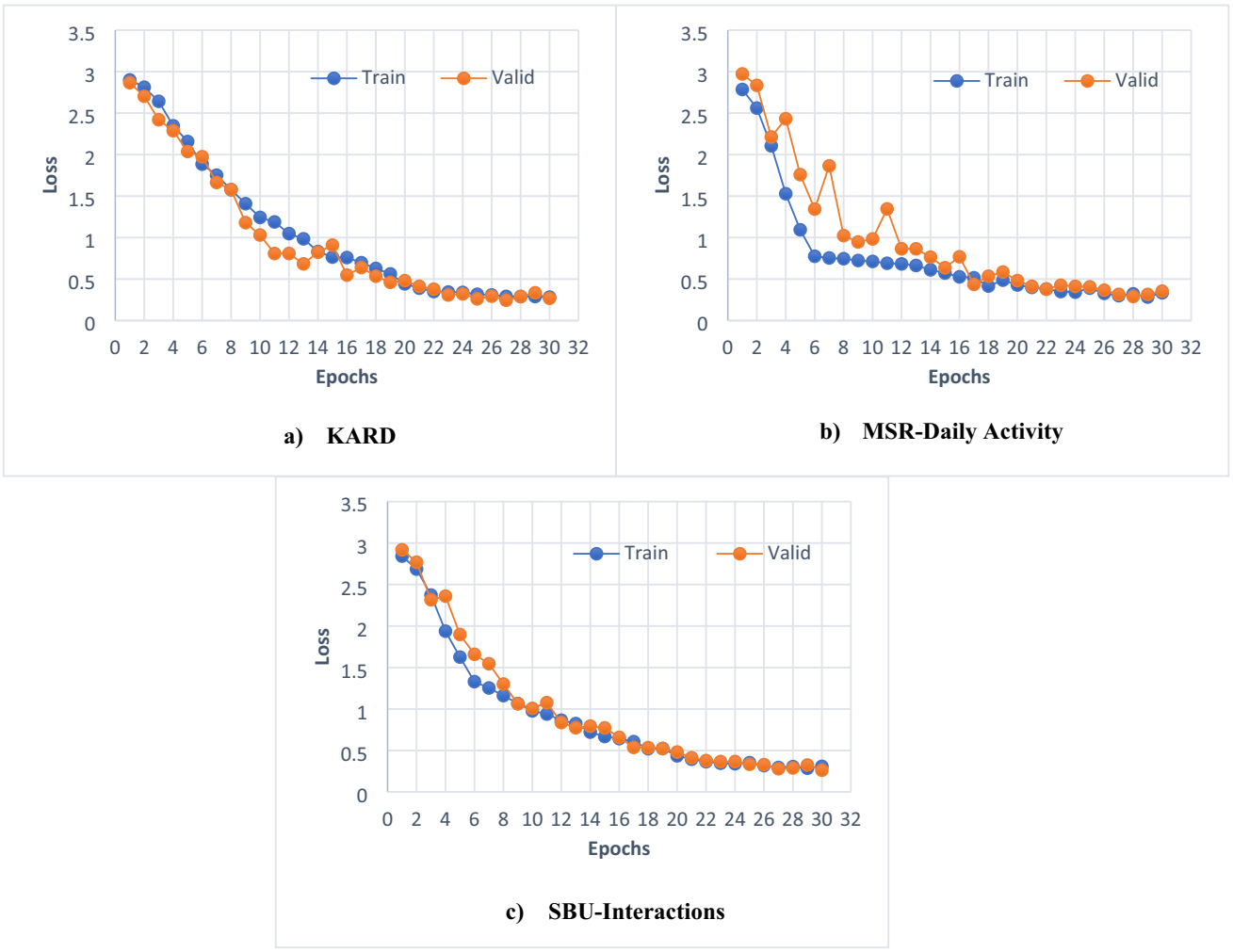


Fig. 10 Training of the proposed architecture over the three datasets using categorical cross entropy loss function

4. **False Positives (FP):** For a particular category, if ground truth is not present but the classifier anticipates its presence.
5. **False Negatives (FN):** For a particular category, if ground truth is present but classifier is not able to identify the same.
6. **Accuracy:** It is simply identified as the ratio of correctly sampled predictions to total observations presents in the dataset i.e., summation of True positives and negatives to summation of all elements present in the confusion matrix (Eq. 17):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{17}$$

7. **Precision:** it identifies how many positive predictions were actually correct i.e., fraction of true positive prediction to all positive predictions. For a given activity ***i*** it is given by (Eq. 18).

$$\text{Precision} = \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i} \tag{18}$$

8. **Recall:** This metric identifies how many actual positive samples were identified correctly i.e., fraction of true positives to all positive samples present in the data. For a given activity ***i*** it is given by (Eq. 19).

$$\text{Recall} = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \tag{19}$$

Using the described metrics, we first evaluate performance for each dataset separately, as observed in Tables 2, 3, and 4. For each dataset, the corresponding confusion matrix is shown in Fig. 11. The results on SBU interaction and KARD are classified to incredible accuracy with state-of-the-art results on MSR-Daily activity. For each dataset's activity classification, we use their corresponding cross-validation

Table 2 Accuracy of the proposed architecture over the KARD dataset

Activity	Precision	Recall
Horizontal arm wave (K1)	1	1
High arm wave (K2)	1	1
Two hand wave (K3)	1	1
Catch cap (K4)	1	0.967
High throw (K5)	1	1
Draw X (K6)	1	1
Draw Tick (K7)	1	1
Toss paper (K8)	1	1
Forward Kick (K9)	1	1
Side Kick (K10)	1	1
Take umbrella (K11)	1	0.867
Bend (K12)	1	1
Hand clap (K13)	1	1
Walk (K14)	0.857	1
Phone call (K15)	1	1
Drink (K16)	1	1
Sit down (K17)	1	1
Stand up (K18)	1	1
Total accuracy	99.07%	

Table 3 Accuracy of the proposed architecture over the MSR-daily activity dataset

Activity	Precision	Recall
Drink (M1)	1	0.95
Eat (M2)	1	1
Read book (M3)	1	0.95
Call cellphone (M4)	1	1
Write on paper (M5)	1	1
Use laptop (M6)	0.952	1
Use vacuum cleaner (M7)	1	1
Cheer up (M8)	1	1
Sit still (M9)	1	0.95
Toss paper (M10)	0.952	1
Play Game (M11)	1	1
Lie down on sofa (M12)	1	1
Walk (M13)	1	1
Play guitar (M14)	1	1
Stand up (M15)	0.95	0.95
Sit down (M16)	0.952	1
Total accuracy	98.75%	

Table 4 Accuracy of the proposed architecture over the SBU-Interactions dataset

Activity	Precision	Recall
Approaching (S1)	0.989	0.98
Departing (S2)	0.989	0.95
Kicking (S3)	1	1
Punching (S4)	1	1
Pushing (S5)	1	1
Hugging (S6)	0.98	1
Shaking hands (S7)	0.99	1
Exchanging (S8)	0.98	1
Total accuracy	99.13%	

techniques suggested by the respective author. MSR-Daily Activity dataset applies half of the human subjects in each activity for training and the other half for testing. In KARD, the entire dataset is distributed into three different activity sets, each containing eight different activities, the proposed algorithm is then tested for each set separately. A human subject-dependent five-fold cross-validation is used for the SBU-interactions dataset.

From the confusion matrices, it is clear that similar activities are being incorrectly classified, but that too in less frequency. The activities 'take umbrella' and 'catch-cap' are wrongly classified as 'walk' for KARD due to the presence of slight walk movements in these classes. For the MSR-Daily activity, the similarity of actions between 'reading'- 'using laptop', 'standing-up'- 'sitting-still' and 'stand-up'- 'paper-toss' can be again attributed to slight movement similarities between the given activities. Similarly, SBU-interaction datasets confusions are similar in nature between activities: 'Approaching,' 'Departing,' 'Hugging,' 'Exchanging,' and 'Shaking-Hands.' Despite these confusions, the misclassifications are very small, which makes the proposed technique useful for practical deployment.

Comparing the results of our proposed technique with other state-of-the-art methods, our proposed methodology outperforms all other methods over the three activity recognition datasets. As observed from Table 5, the results over SBU-interaction and KARD have over 99% accuracy. Most of the compared methods in this domain used skeleton maps of the humans for devicing their activity classes [36, 39–41], rely solely on the skeleton maps i.e., posture involved in the video, [42] on the other hand utilized the depth map information using R-transform and zernike moments in just the depth video. The accuracy on MSR-Daily Activity falls

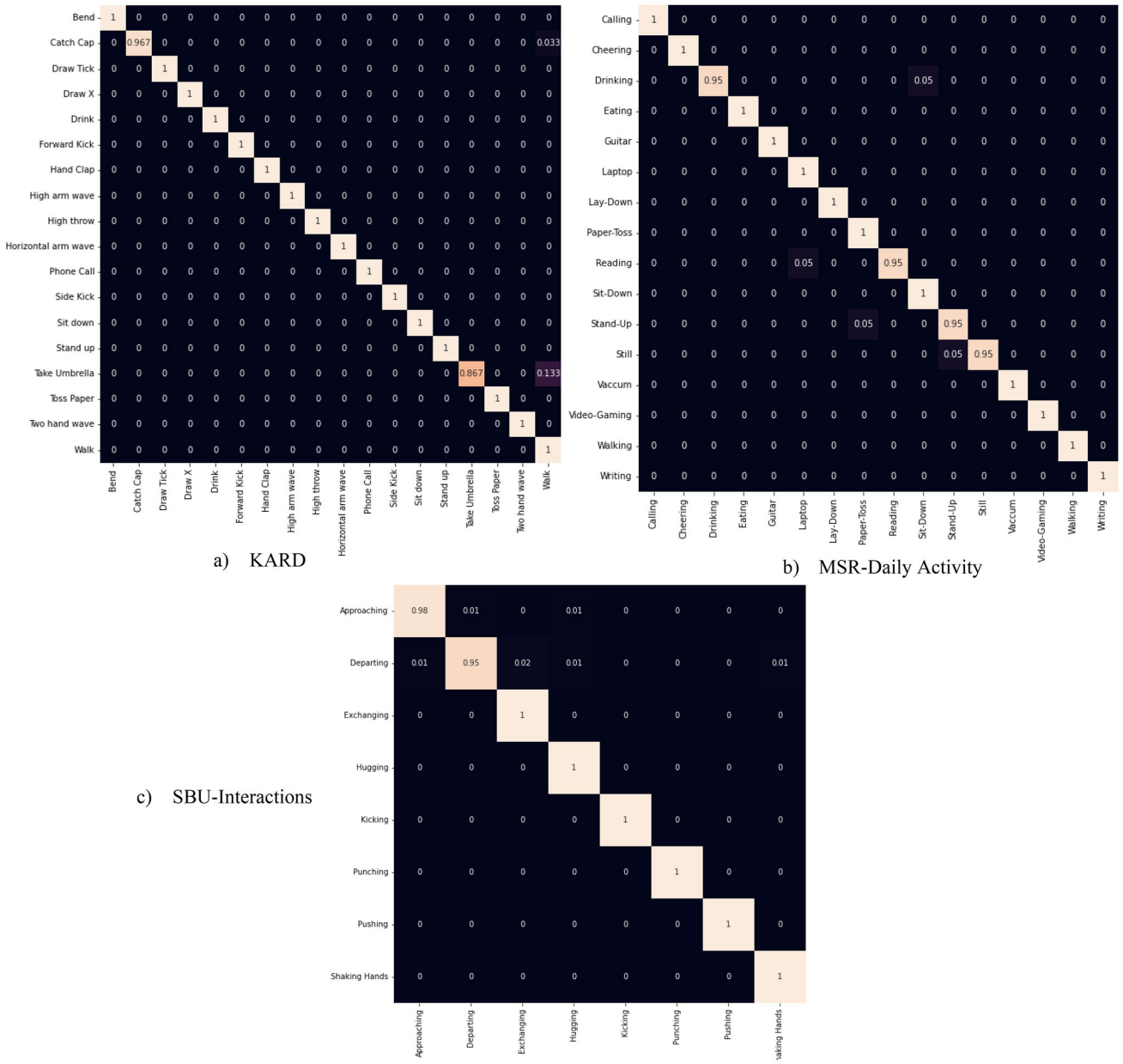


Fig. 11 Confusion matrices for the proposed architecture over the three datasets (zoom in for better view)

short of 99% but still outperforms all the other techniques compared with. For MSR-Daily dataset, [20, 43, 44] all use multimodal input, including depth maps along with RGB frames for coupled prediction, whereas [45] again used skeletal information with [46] using 3D convolution to learn the action dependencies. Even without the pre-processed video frames, the proposed architecture produced some comparable results, showing the efficiency of the implemented Conv-LSTM architecture. Further, the better results over the MSR-Daily Activity can be attributed to the complexity of the foreground, which is reduced by our proposed pipeline. The focus on the human subject plays a vital role in making the Conv-LSTM architecture quickly identify the concurrent human Activity. Similar results are observed over SBU Interaction dataset, achieving a 99% accuracy over the previous state of the art methods. Again, [20, 47, 48] apply multimodal fusion with either skeletal or depth data throughout the frames aiding the models to map for the activity, with [19, 38] using graph based modeling to learn the occuring activity.

The method, though accurate, needs to be improved for longer context windows and for far more complex backgrounds. The datasets tested upon have relatively easier setting to test upon. Further, as we move toward a longer

Table 5 Accuracy of the proposed architecture compared with state-of-the-art techniques

Dataset	Technique	Accuracy	#Parameters
KARD	Gaglio et al. [36]	94.5	–
	Dhiman et al. [42]	96.64	26 M
	Ashwini et al. [39]	97.22	34 M
	Madany et al. [40]	98.5	17 M
	Cippitelli et al. [41]	98.2	–
	Proposed Model (w/o)	**94.21**	**12 M**
	(With pre-processing)	**99.07**	**19 M**
MSR-daily activity	Singh et al. [20]	94.37	40 M
	Andrade-Ambriz et al. [43]	95.6	–
	Huynh-The et al. [45]	90.63	–
	Zhu et al. [46]	93	36 M
	Shahroudy et al. [44]	97.5	27 M
	Proposed Model (w/o)	**91.47**	**12 M**
	(With pre-processing)	**98.75**	**19 M**
SBU interactions	Yun et al. [38]	87.3	–
	Song et al. [47]	91.51	17 M
	Elboushaki et al. [48]	97.51	–
	Singh et al. [20]	98.7	40 M
	Yan et al. [19]	98.9	–
	Proposed Model (w/o)	**91.78**	**12 M**
	(With pre-processing)	**99.13**	**19 M**

Bold value indicates the accuracy achieved by the proposed model

context length temporal consistency in the segmentation is required to play a vital part in the pre-processing step, which the current proposition lacks in. Adding on to the edge cases, the method might fail in case the activity recognition in a frame constitutes of some '*x*' people in the frame performing an activity with *y* people in the background, the model needs to focus on the *x* people to give the inference on the activity. This could be improved via a tempoal RCNN like method to identify the region where the activity might be happening and include in the pre-processing step. To improve the same we can consider Segmenting objects from relational visual data i.e., human interaction with objects [49, 50] for region identification that corresponds to the activity being performed.

5 Conclusion and future work

In this study, we have presented a robust end-to-end deep learning pipeline for human activity recognition (HAR) in sequential RGB video frames. Our proposed approach integrates a U-net-like segmentation model for effective pre-processing, enhancing the input frames' representation. The segmentation maps generated by this model are then utilized as inputs to a pre-trained Efficient-Net architecture, followed by a Bi-directional Long Short-Term Memory (LSTM) network for learning temporal dependencies and classifying specific activities. The experimental results across three datasets—KARD, MSR Daily Activity, and SBU-Interactions—demonstrate the superiority of our methodology, achieving benchmark accuracy scores exceeding 98 and 99% for MSR Daily Activity and KARD and SBU-Interactions, respectively.

While our proposed method provides a competitive solution for activity recognition, acknowledging its strengths, several limitations and areas for improvement have been identified. The model's performance could be enhanced for longer context windows, particularly in scenarios with more complex backgrounds. Future iterations should focus on improving temporal consistency in the segmentation step, especially for extended temporal contexts. Additionally, addressing edge cases, such as scenarios with multiple individuals in the frame engaging in different activities, requires a more sophisticated approach, potentially involving temporal Region-based Convolutional Neural Networks (RCNN) for precise region identification.

Looking forward, we propose future works to explore more advanced architectures, including the integration of transformers for sequential modeling, which could further enhance the network's ability to capture long-range dependencies in video data. Extending the segmentation model to identify and focus on objects that humans interact with could significantly improve the recognition of complex activities, particularly in sports scenarios. Incorporating relational visual data and human-object interaction studies may provide valuable insights for region identification corresponding to various activities.

In conclusion, our study lays a solid foundation for effective human activity recognition through innovative integration of Conv-LSTM networks and segmentation-based pre-processing. The identified limitations and proposed future directions open avenues for continued research, promising advancements in the field of activity recognition and computer vision.

Data availability statement Datasets used in this work are available in public repository. https://data.mendeley.com/datasets/k28dtm7tr6/1, https://sites.google.com/view/wanqingli/data-sets/msr-dailyactivity3d, https://cove.thecvf.com/datasets/57#:~:text=A%20complex%

20human%20activity%20dataset,depth%20and%20motion%20capture%20data.

Declaration

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. Karpathy, A., Toderici, G., Shetty S., Leung, T., Sukthankar, R., Li, F. F.: Large-scale video classification with convolutional neural networks. In: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 1725–1732 (2014). https://doi.org/10.1109/CVPR.2014.223.
2. Zeng, M., et al.: Convolutional Neural Networks for human activity recognition using mobile sensors. In: Proc. 2014 6th Int. Conf. Mob. Comput. Appl. Serv. MobiCASE 2014, vol. 6, pp. 197–205, (2015). https://doi.org/10.4108/icst.mobicase.2014.257786.
3. Dhiman, C., Vishwakarma, D. K.: A review of state-of-the-art techniques for abnormal human activity recognition. Eng. Appl. Artif. Intell., **77**(June 2018), 21–45 (2019). https://doi.org/10.1016/j.engappai.2018.08.014.
4. Dhiman, C., Vishwakarma, D. K., Agarwal, P.: Part-wise spatiooral attention driven CNN-based 3D human action recognition. ACM Trans. Multimed. Comput. Commun. Appl., **17**(3) (2021). https://doi.org/10.1145/3441628.
5. Dhiman, C., Vishwakarma, D. K.: View-invariant deep architecture for human action recognition using two-stream motion and shape temporal dynamics. IEEE Trans. Image Process. **29**, 3835–3844 (2020). https://doi.org/10.1109/TIP.2020.2965299.
6. Vishwakarma, D.K., Kapoor, R.: Hybrid classifier based human activity recognition using the silhouette and cells. Expert Syst. Appl. **42**(20), 6957–6965 (2015). https://doi.org/10.1016/j.eswa.2015.04.039
7. Straka, M., Hauswiesner, S., Rüther, M., Bischof, H.: Skeletal graph based human pose estimation in real-time. In: BMVC,: Proc. Br. Mach. Vis. Conf. **2011**, 2011 (2011). https://doi.org/10.5244/C25.69
8. Simonyan, K., Zisserman, A.: Two-stream convolutional networks for action recognition in videos. Adv. Neural. Inf. Process. Syst. **1**(January), 568–576 (2014)
9. Jain, A., Vishwakarma, D. K.: State-of-the-arts violence detection using ConvNets. In: Proc. 2020 IEEE Int. Conf. Commun. Signal Process. ICCSP 2020, pp. 813–817, (2020). https://doi.org/10.1109/ICCSP48568.2020.9182433.
10. Yadav, A., Vishwakarma, D.K.: A deep learning architecture of RA-DLNet for visual sentiment analysis. Multimed. Syst. **26**(4), 431–451 (2020). https://doi.org/10.1007/s00530-020-00656-7
11. Sabih, M., Vishwakarma, D. K.: Crowd anomaly detection with LSTMs using optical features and domain knowledge for improved inferring. Vis. Comput., no. 0123456789, (2021). https://doi.org/10.1007/s00371-021-02100-x.
12. Jeevan, M., Jain, N., Hanmandlu, M., Chetty, G.: Gait recognition based on gait pal and pal entropy image. In: 2013 IEEE Int. Conf. Image Process. ICIP 2013 - Proc., pp. 4195–4199 (2013). https://doi.org/10.1109/ICIP.2013.6738864.
13. Huang, W., Zhang, L., Wu, H., Min, F., Song, A.: Channel-equalization-HAR: a light-weight convolutional neural network for wearable sensor based human activity recognition. IEEE Trans. Mob. Comput. **22**(9), 5064–5077 (2023). https://doi.org/10.1109/TMC.2022.3174816
14. Wenbo, H., Zhang, L., Wang, S., Wu, H., Song, A.: Deep ensemble learning for human activity recognition using wearable sensors via filter activation. ACM Trans. Embed. Comput. Syst. **22**(1), 1–23 (2022). https://doi.org/10.1145/3551486
15. Garcia, N. C., Morerio, P., Murino, V.: Modality distillation with multiple stream networks for action recognition. In: Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11212 LNCS, pp. 106–121 (2018). https://doi.org/10.1007/978-3-030-01237-3_7.
16. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. In: 32nd AAAI Conf. Artif. Intell. AAAI **2018**, 7444–7452 (2018)
17. Ji, S., Xu, W., Yang, M., Yu, K.: 3D Convolutional neural networks for human action recognition. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 221–231 (2013). https://doi.org/10.1109/TPAMI.2012.59
18. Baccouche, M., Mamalet, F., Wolf, C., Garcia, C., Baskurt, A.: Sequential deep learning for human action recognition. In: Proc. Second Int. Conf. Hum. Behav. Underst., vol. 34, no. 8, 2012.
19. Yan, G., Hua, M., Zhong, Z.: Multi-derivative physical and geometric convolutional embedding networks for skeleton-based action recognition[Formula presented]. Comput. Aided Geom. Des., **86** (2021). https://doi.org/10.1016/j.cagd.2021.101964.
20. Singh, T., Vishwakarma, D.K.: A deeply coupled ConvNet for human activity recognition using dynamic and RGB images. Neural Comput. Appl. **33**(1), 469–485 (2021). https://doi.org/10.1007/s00521-020-05018-y
21. Zebhi, S., AlModarresi, S.M.T., Abootalebi, V.: Human activity recognition using pre-trained network with informative templates. Int. J. Mach. Learn. Cybern. **12**(12), 3449–3461 (2021). https://doi.org/10.1007/s13042-021-01383-9
22. Fei-Fei, L., Deng, J., Li, K.: ImageNet: constructing a large-scale image database. J. Vis. **9**(8), 1037–1037 (2010). https://doi.org/10.1167/9.8.1037
23. Ullah, A., Ahmad, J., Muhammad, K., Sajjad, M., Baik, S. W.: Action recognition in video sequences using deep bi-directional LSTM with CNN features. IEEE Access, **6**, 1155–1166 (2017). https://doi.org/10.1109/ACCESS.2017.2778011.
24. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q.: Densely connected convolutional networks. In: Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017, pp. 2261–2269 (2017). https://doi.org/10.1109/CVPR.2017.243.
25. Weng, W., Zhu, X.: UNet: convolutional networks for biomedical image segmentation. IEEE Access **9**, 16591–16603 (2015). https://doi.org/10.1109/ACCESS.2021.3053408
26. "Supervisely Person Dataset - Datasets - Supervisely." https://supervise.ly/explore/projects/supervisely-person-dataset-23304/datasets (accessed May 30, 2021).
27. Tan, M., Le, Q. V.: EfficientNet: Rethinking model scaling for convolutional neural networks. In: 36th Int. Conf. Mach. Learn. ICML 2019, vol. 2019-June, pp. 10691–10700 (2019).
28. He, K., Zhang, S., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-December, pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90.
29. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-December, pp. 2818–2826 (2016). https://doi.org/10.1109/CVPR.2016.308.
30. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-January, pp. 1800–1807 (2017). https://doi.org/10.1109/CVPR.2017.195.

31. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-ResNet and the impact of residual connections on learning. In: 31st AAAI Conf. Artif. Intell. AAAI **2017**, 4278–4284 (2017)
32. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L. C.: MobileNetV2: inverted residuals and linear bottlenecks. In: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 4510–4520 (2018). https://doi.org/10.1109/CVPR.2018.00474.
33. Hu, J.: Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 7132–7141 (2018). https://doi.org/10.1109/CVPR.2018.00745.
34. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **1780**, 1735–1780, 1997, [Online]. https://doi.org/10.1162/neco.1997.9.8.1735.
35. Schuster, M., Paliwal, K. K.: Bidirectional recurrent neural networks. IEEE Trans. Signal Process. **45**(11), 2673–2681 (1997). [Online]. Available: https://doi.org/10.1109/78.650093.
36. Gaglio, S., Lo Re, G., Member, S., Morana, M.: Human activity recognition process using 3-D posture data. IEEE Trans. Human-Mach. Syst. **45**(5), 586–597 (2015). https://doi.org/10.1109/THMS.2014.2377111.
37. Wang, J., Liu, Z., Wu, Y., Yuan, J.: Mining actionlet ensemble for action recognition with depth cameras. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2012, pp. 1290–1297 (2012). https://doi.org/10.1109/CVPR.2012.6247813.
38. Yun, K., Honorio, J., Chattopadhyay, D., Berg, T. L., Samaras, D.: Two-person interaction detection using body-pose features and multiple instance learning. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2012, pp. 28–35 (2012). https://doi.org/10.1109/CVPRW.2012.6239234.
39. Ashwini, K., Amutha, R.: Compressive sensing based recognition of human upper limb motions with kinect skeletal data. Multimed. Tools Appl. **80**(7), 10839–10857 (2021). https://doi.org/10.1007/s11042-020-10327-4
40. El Madany, N.E.D., He, Y., Guan, L.: Integrating entropy skeleton motion maps and convolutional neural networks for human action recognition. In: Proc. IEEE Int. Conf. Multimed. Expo (2018). https://doi.org/10.1109/ICME.2018.8486480
41. Cippitelli, E., Gasparrini, S., Gambi, E., Spinsante, S.: A human activity recognition system using skeleton data from RGBD sensors. Comput. Intell. Neurosci. (2016). https://doi.org/10.1155/2016/4351435.
42. Dhiman, C., Vishwakarma, D.K.: A robust framework for abnormal human action recognition using\boldsymbol{\mathcal{r}}-transform and zernike moments in depth videos. IEEE Sens. J. **19**(13), 5195–5203 (2019). https://doi.org/10.1109/JSEN.2019.2903645
43. Andrade-Ambriz, Y.A., Ledesma, S., Ibarra-Manzano, M.A., Oros-Flores, M.I., Almanza-Ojeda, D.L.: Human activity recognition using temporal convolutional neural network architecture. Expert Syst. Appl. **191**(December), 2022 (2021). https://doi.org/10.1016/j.eswa.2021.116287
44. Shahroudy, A., Ng, T.T., Gong, Y., Wang, G.: Deep multimodal feature analysis for action recognition in RGB+D videos. IEEE Trans. Pattern Anal. Mach. Intell. **40**(5), 1045–1058 (2018). https://doi.org/10.1109/TPAMI.2017.2691321
45. Huynh-The, T., et al.: Hierarchical topic modeling with pose-transition feature for action recognition using 3D skeleton data. Inf. Sci. (Ny) **444**, 20–35 (2018). https://doi.org/10.1016/j.ins.2018.02.042
46. Zhu, J. et al.: Action machine: rethinking action recognition in trimmed videos. (2018). [Online]. Available: http://arxiv.org/abs/1812.05770.
47. Song, S., Lan, C., Xing, J., Zeng, W., Liu, J.: An end-to-end spatio-temporal attention model for human action recognition from skeleton data. In: 31st AAAI Conf. Artif. Intell. AAAI, 2017, 4263–4270 (2017)
48. Elboushaki, A., Hannane, R., Afdel, K., Koutti, L.: MultiD-CNN: A multi-dimensional feature learning approach based on deep convolutional networks for gesture recognition in RGB-D image sequences. Expert Syst. Appl. **139**, 112829 (2020). https://doi.org/10.1016/j.eswa.2019.112829
49. Kim, D. J., Sun, X., Choi, J., Lin, S., Kweon, I. S.: Detecting human-object interactions with action co-occurrence priors. In: Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 12366 LNCS, pp. 718–736 (2020). https://doi.org/10.1007/978-3-030-58589-1_43.
50. Lu, X., Wang, W., Shen, J., Crandall, D.J., Van Gool, L.: Segmenting objects from relational visual data. IEEE Trans. Pattern Anal. Mach. Intell. **44**(11), 7885–7897 (2022). https://doi.org/10.1109/TPAMI.2021.3115815

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.



Shaurya Gupta is pursuing a Bachelor of Technology in Engineering from Delhi Technological University, Delhi. His current research interest includes Object Recognition, Crowd Nanmlies Detection, Human Pose estimation, and Deep learning.



Dinesh Kumar Vishwakarma (M'16, SM'19) received Doctor of Philosophy (Ph.D.) degree in the field of Computer Vision from Delhi Technological University (Formerly Delhi College of Engineering), New Delhi, India, in 2016. He is currently a Professor with the Department of Information Technology, Delhi Technological University, New Delhi. His current research interests include Computer Vision, Deep Learning, Sentiment Analysis, Fake News Detection, Crowd Analysis, Human Action, and Activity Recognition. He is a reviewer of various Journals/Transactions of IEEE, Elsevier, and Springer. He has been awarded a research excellence award in the years 2017, 2018, 2019, and 2020 for his outstanding contribution in the research.



Nitin Kumar Puri received Doctor of Philosophy (Ph.D.) degree in from Punjab University, Chandidarh. He is currently a Professor with the Department of Applied Physics, Delhi Technological University, New Delhi. His current research interests include Sensors, Bio-Sensors, Nanotechnology, Ion-Beam modification of materials, and Thin films.

Archives of Computational Methods in Engineering
https://doi.org/10.1007/s11831-024-10080-z

REVIEW ARTICLE

A Systematic and Comprehensive Review on 2-D and 3-D Numerical Modelling of Stirling Engine

Vaibhav Singh[1] · **Anil Kumar**[1,2]

Received: 15 October 2023 / Accepted: 17 January 2024
© The Author(s) under exclusive licence to International Center for Numerical Methods in Engineering (CIMNE) 2024

Abstract
The present study summarises a comprehensive review of the simulation modelling of the Stirling engine. A Stirling engine, characterized as an external combustion engine, is versatile enough to operate on various fuels, including solar energy and waste heat. Among the different analysis approaches for the Stirling Engine, fourth-order analysis, which is Computational Fluid Dynamics (CFD) analysis, stands out as the sole method that takes into account the actual geometry of the engine. It is capable of identifying intricate gas flow patterns, as well as non-uniform pressure and temperature distributions within the engine. This results in better understanding and optimisation of the design parameters of the engine. The primary goal of this study is to compile and summarize 2D and 3D CFD simulation investigations, categorizing them based on the software employed, including ANSYS Fluent, OpenFOAM, and COMSOL. The focus will be on delineating the study type, computational domain, and the eddy viscosity models utilized for the modeling of Stirling engines. Comparison of various models and their deviation from experimental results is also listed. It was found that considering the radiation model resulted in an improvement of accuracy by 11%. The impinging effect, noted by CFD analysis, is responsible for vortex formation that improved heat transfer, resulting in a discrepancy of over 80% between the prediction of power output by the second-order method and CFD. For accuracy, simulation in full domain 3D is preferred over partial domain 3D and 2D CFD studies. However, it will require more computational resources, accurate 3D model and boundary conditions.

1 Introduction

The world is currently experiencing an energy and resource shortage. To protect scientific and technical progress, numerous organisations have been actively discovering and examining alternate energy sources [1, 2]. A good energy mix of renewable and non-renewable sources is the need of the hour. For this several hybrid systems like combined heat and power systems or grid-connected renewable systems with or without energy storage are being investigated [2, 3].

For more than two centuries, the Stirling engine has been in use, proving versatile in various applications. Since they only have a few moving parts, they have a high efficiency of roughly 35%, little vibration and noise, and straightforward construction. Additionally, Stirling engines can run on environmentally beneficial energy sources like biomass, waste heat, or solar energy [4, 5].

Stirling engines based on solar energy and hybrid systems were investigated by many authors [6–12]. Popular applications of solar-based Stirling engine systems include Solar Stirling Electric Power Generation, Combined Heat and Power, Off Grid Electrification, Hybridisation and Storage, Water distillation and desalination, Water pumping, etc.

Alpha, beta, and gamma-type Stirling engines are the three different types of Stirling engines [13]. Free-piston Stirling engines are those that operate without a piston drive mechanism [14].

One of the energy systems that needs a thorough thermodynamic analysis for its operation to be understood is the Stirling engine [15]. Such a study frequently calls for a complex model. Stirling engines are designed and their performance is predicted using five different techniques: zero-order analysis, first-order analysis, second-order analysis, third-order analysis (also known as the nodal design approach), and fourth-order analysis (also known as CFD simulation). In this context, the term "order" refers to

✉ Anil Kumar
anilkumar76@dtu.ac.in

[1] Department of Mechanical Engineering, Delhi Technological University, Delhi 110 042, India

[2] Centre for Energy and Environment, Delhi Technological University, Delhi 110 042, India

Published online: 20 February 2024

classification based on the model's level of complexity, with a higher order indicating a more complex model [15, 16].

The first-order analysis ignores all losses. The majority of the irreversibilities, which are exhibited during the operation of Stirling engines, are not taken into account, making this form of study highly straightforward yet unreliable. Using Baele number or the slightly advanced method by Iwamoto, which includes experiment results for performance prediction, are examples of first-order analysis [17, 18]. In second-order models, computations initially proceed under the assumption of no losses. After achieving convergence in the solution, various types of losses are then independently assessed. These encompass pressure drop, losses in the appendix gap, conduction and radiation losses, heat exchanger losses, as well as inefficiencies in regenerator thermal inefficiency [19, 20]. These losses include pressure drop, heat exchanger losses, appendix gap losses, conduction and radiation losses, regenerator thermal inefficiency losses, etc. A second-order model known as PSVL (polytropic analysis of Stirling engine with varied losses) was presented by Babaelahi and Sayyaadi [19]. This model replaced the early model's adiabatic compression and expansion processes with polytropic ones. Yang et al. recently introduced an enhanced second-order adiabatic model called the Incorporated Pressure Drop-modified Simple Model (IPD-MSM), building upon the Simple model. It showed better performance prediction at high pressure and frequency conditions as compared to other adiabatic models like Combined Adiabatic-Finite Speed (CAFS) and Improved Simple Analytical Model (ISAM) [21]. In third-order methods, the engine's geometry undergoes discretization into a network of nodes and control volumes, followed by the solution of a set of differential equations [22]. The examination of third-order models is generally intricate, demanding increased computational effort and time compared to second-order models [23]. Using third-order analysis, Aksoy et al. [24] divided the Stirling engine's total volume into 22 nodal volumes in order to compare and contrast the operation of the crank drive and rhombic drive mechanisms.

The fourth-order analysis is the CFD analysis, in the present study; a review of major CFD studies for the Stirling engine has been presented. It is capable of recognising complex gas flow behaviour, uneven pressure and temperature distributions in the engine and provides better understanding and optimisation of the design parameters of the engine. It also enables the identification of impinging effect and formation of vortices, which results in the enhancement of heat transfer within the engine. CFD studies have been categorized as either 2D-based or 3D-based and presented under the headings of the software on which they have been conducted. Tables summarizing key features and remarks about the studies have also been presented. Finally, a table comparing deviation in indicated power, from the simulation result to experimental values, has been presented which highlights the eddy viscosity models used and its computational domain. Equations used in these models are also presented. Any Stirling engine type can use the CFD approach because any variation of a parameter is already included in the Navier-Stokes equations. Most accurate analysis can be done by CFD but it also requires the most inputs like an accurate 3D model, and boundary conditions, the selection of appropriate eddy viscosity models and is computationally expensive, i.e. it takes a lot of time for a convergent solution.

2 Simulation Modelling of Stirling Engine

ANSYS Fluent is a general-purpose industry-leading CFD programme used to model fluid flow, heat and mass transfer, chemical processes, and more. It is renowned for its extensive physics modelling capabilities and industry-leading accuracy [25]. COMSOL Multiphysics, a software package encompassing finite element analysis, solvers, and simulation tools, finds application across diverse realms of physics and engineering. It excels in handling coupled phenomena and Multiphysics scenarios, supporting systems of partial differential equations and providing user-friendly interfaces based on standard physics principles [26]. .

OpenFOAM, an open-source C++toolbox abbreviated from "Open-source Field Operation And Manipulation," serves the purpose of developing tailored numerical solvers, as well as pre and post-processing tools for problems in continuum mechanics. Its notable application lies in tackling CFD challenges [27]. .

2.1 2D-Based Studies

2.1.1 Custom Codes

Mahkamov [28], in 2006, conducted a two-dimensional axisymmetric computational fluid dynamic study for the working process of an alpha-type Solar Stirling engine. The standard k-ε turbulence model was used, and results were validated which indicated an error of approximately 40% in power prediction. Even though the error is huge, it was significantly better than the power predicted by lower-order methods. This is attributed to the fact that CFD represents the complexities of fluid flow and temperature variation better than second-order methods. According to Mahkamov, for layouts other than alpha, the axisymmetric calculation scheme will provide only a preliminary estimation of the engine performance. Three-dimensional CFD is preferred for accurate results.

Salazar and Chen [29] used their own compressible flow CFD code to research a straightforward beta-type engine. Because of the low engine speed, the fluid flow

was characterized as laminar during the construction of the computer model within an axisymmetric and 2D domain. The engine lacked external regeneration and exhibited low power. The observation was made that flow impingement significantly influenced heat transfer, resulting in a notably non-uniform temperature distribution throughout the engine at any given moment.

2.1.2 COMSOL

Alfarawi et al. [30]used CFD simulation to investigate how the dead volume and phase angle affected a gamma-type engine. The flow of fluid was considered laminar, and the Lagrangian-Eulerian (ALE) technique was used to manage moving engine components and changing cylinder volume. One of the most effective strategies to regulate Stirling engine output is considered to be changing phase angle. Phase angle affects three key engine parameters: pressure amplitude, total volumetric change of gas, and heat transfer, all of which have a direct impact on the engine's indicated power.

2.1.3 Open FOAM

Della Torre et al. [31] conducted a simulation of a 2D axisymmetric beta-type Stirling engine in their study, utilizing the OpenFOAM open-source software and adopting an experimental model developed by the University of Valenciennes in France. The engine was first parametrically analysed before simulation findings were assessed using experimental data. In the parametric study, the effects of engine thermal efficiency and engine power output were assessed for the kind of average pressure of engine, heat received by the engine, working fluid and regenerator characteristics.

OpenFOAM, an open-source programme, was used to model a beta-type Stirling engine by Mohammadi and Jafarian [32]. The Ground Power Unit-3 (GPU 3) Stirling engine's CFD model with laminar flow produces results that are lower than 1% different from those produced by the k-ε Shear Stress Transport(SST) turbulence model, indicating that laminar flow is predominant. The average pressure is 1.5 bars, and the power output is 13.2 W. The 2D representation of the intricate engine shape oversimplified the model, making several assumptions that neglected factors such as shuttle effects, mass leakages, mechanical friction, and regenerator inefficiency. As a result, the actual flow characteristics of the regenerator and the mechanism of the heat transfer intricacies mechanism between fluid and solid were disregarded.

2.1.4 ANSYS Fluent

Ben Mansour et al. [16] conducted CFD study of a Stirling engine by considering internal radiation, which is neglected by most researchers. The charge pressure of the engine was atmospheric with no external regenerator and laminar flow is assumed. The largest discrepancy between the CFD model's predictions and the experimental data was 7%. A 13% underestimating of power production was reported because of absence of radiation heat transfer. The inaccuracy in estimating power output reached 18% when the cumulative effects of radiation were disregarded. The Discrete Ordinates (DO) radiation model was found to provide results that are more accurate and less computationally expensive than the surface-to-surface (S2S) radiation model. The working fluid can also have an impact on internal radiation heat transfer.

In a numerical study, Abuelyamen and Ben-Mansour [33] compared the energy efficiency and power output of three different types of Stirling engines (α, β and γ). Internal radiation heat transmission within the Stirling engine was taken into account. The three Stirling engine variants did not include regenerators. A set of experimental data from the literature was used to validate the model. The annular connecting pipe was added to the α-type Stirling engine in place of the circular pipe because the narrow channel results in a high heat transfer rate. After that, there was an increase in power output (from 0.9 to 9.856 W) and thermal efficiency (from 1.8 to 10.5%). It was found that vortices are the responsible for good mixing of gas, which in turn leads to lower cold space temperature, which reduces the compression work, hence increasing the power output and efficiency. 2D CFD studies performed on Stirling engines are summarised in Table 1.

2.2 3D-Based Studies

2.2.1 Custom

Chen et al. [34] created a proprietary 3-dimensional CFD code in-house to model the heat transfer in a quarter of the computational domain of a basic low-temperature difference gamma-type Stirling engine. Mechanical friction and thermal radiation were not considered. There was no external regenerator and the power produced is low (0.62 W) with low efficiency (4.82%), primarily due to low engine speed and low-temperature difference b/w hot and cold ends. The study concluded that impingement is the primary mechanism of heat transfer in both the compression and expansion chambers, leading to a non-uniform gas temperature distribution within the engine spaces. Consequently, it is not advisable to rely on the application of constant heat transfer coefficients and the assumption of uniform temperature in the chambers.

Table 1 Summary of 2D CFD studies performed on Stirling Engines

Author	Year	Engine Type	Model used	Domain	Software used	Remarks
Mahkamov [28]	2006	Alpha	Standard k-ε turbulence	Axis symmetric	Custom	• Over prediction by 0.5 kW, approx. 40% deviation. • Heat flux produced by the parabolic concentrator is not uniform on the receiver surface.
Salazar, Chen [29]	2014	Beta		Axis symmetric	Custom	• The assumption of uniform temperatures of chamber and heat transfer coefficients as constant, to address heat transfer is overly simplistic and fails to accurately represent reality.
Alfarawi et al. [30]	2016	Gamma	Laminar	Planar	COMSOL	• Analysed the influence of dead volume and phase angle. • Max deviation in indicated power was 9%.
Della Torre [31]	2014	Beta	k-ω	Axis Symmetric	OpenFOAM	• The ideal cycle is significantly • Larger than the real cycle, on a PV graph. • Power output of the real cycle is 22% of the ideal cycle
Mohammadi [32]	2018	Beta	k-ω SST turbulence	Axis Symmetric	OpenFOAM	• 15.1% difference as compared to experimental results.
Ben-Mansour et al. [16]	2017	Beta	Laminar	Axis Symmetric	Ansys Fluent	• Considered internal radiation heat transfer. • Discrete Ordinates (DO) radiation models are better than Surface-to-surface(S2S). • Deviation of 7% with experimental results. • Deviation of 18% when radiation effects in total were ignored.
Abuelyamen and Ben-Mansour [33]	2018	Alpha, Beta & Gamma	Laminar	Axis Symmetric	Ansys Fluent	• Compared performance between alpha, beta and gamma types of Stirling engines, with no external regenerators. • Considered internal radiation heat transfer. • Gamma(9.223 W and 9.8%) has highest Power Output efficiency followed by Beta(8.634 W and 7.5%) and Aplha(0.908 W and 1.8%). Upon modification, Alpha became the highest(9.856 W and 10.5%).

2.2.2 COMSOL

In their study, Almajri et al. [35] utilized COMSOL-Multiphysics software to conduct a 3-dimensional parametric simulation of a standard Stirling converter of alpha-type. They delved into various operating conditions and design elements, including temperatures at both ends, charge pressure, regenerator porosity, metal matrix wire diameter, and geometric factors. The application of the optimized new design parameters and operating conditions resulted in a 65% increase in power output, rising from 146 to 240 W.

2.2.3 ANSYS Fluent

Mahkamov [36], in 2006, performed the first successful 3D CFD simulation of a Gamma-type Stirling engine run by biomass. Compared to 30% for the second-order results, the forecasts were only 12%–18% off from the experimental data. k-ε turbulence model was used and re-meshing was accomplished by removing or adding cell layers above the power piston and displacer piston during the stroke. The results showed that, despite identical shapes, the velocity field varied dramatically between tubes. These variations could result in a highly uneven thermal load on the heater's components and ineffective operation. Additionally, it was shown that while the mechanical efficiency of harmonic engines is greater than that of a discontinuous drive, the thermodynamic efficiency of engines with harmonic piston displacement is lower than that of an ideal Stirling engine. The original engine encountered diminished power output attributed to regenerator's notable pressure drop, the existence of a "dead" zone in the compression space, and gas entrapment in the connecting pipe between the 2 halves of the compression space. To enhance power output, the engine was subsequently converted from gamma to alpha type.

Li et al. [37] using Ansys Fluent modeled a small-size Stirling engine of alpha-type in three dimensions. The assessment of engine performance involved evaluating pressure drop, heat exchange rate in the regenerator, and variations in hot heat exchanger and regenerator geometries. When the findings were compared to experimental data, it became evident that regular-shaped microchannels outperformed wire mesh as a regenerator.

Cheng and Chen [38] analysed the effect of different working parameters on a 1 kW β-type Stirling engine. The parameters included filling pressure, heating temperature, and regenerator porosity. The relative performance of helium, nitrogen and air was also studied and it was found that helium leads to the highest engine performance and nitrogen is better than air. The trend of the numerical predictions agreed with that of the experiment data, but the actual values differed by over 27%. Even though some assumptions were good, the study failed to accurately predict the engine output.

When Xiao et al. [39]used CFD simulation to investigate the operating parameters of a beta-type Stirling engine, the largest divergence was 13.4%. These variables included wire diameter, matrix mesh, regenerator length, and heat exchanger tube diameter. Redesigned heat exchangers resulted in a 40% increase in calculated power output. Some assumptions were not good, which might lead to inaccuracies like simplified geometry, excessive solid parts, thickness layers and regenerator with perfect regeneration. Effects of thermal radiation and mechanical friction were also ignored. The dead volume of heat exchangers was significantly reduced in the optimised model, increasing thermal efficiency and output power with little flow resistance loss. It has been observed that more dead space causes the pressure to phase shift.

In his research, Kuban et al. [40] looked at the effects of the engine speed, initial charge pressure, and heater temperature on the engine's flow, heat transfer, and indicated power characteristics. In their study, dynamic meshing and re-meshing techniques were utilized to clarify the movement of both the compression piston and expansion piston. The outcomes demonstrated favorable alignment with the experimental results. The author suggests that because they have a major impact on the flow field, whole engines should be taken into account for CFD rather than quarter engines. The k-ε turbulence model is realistic. As rotational speed increases, there are less and smaller differences between the k-ω-SST and k-ε realisable models; the k-ε realisable model predicts more. The differences between the k-ω-SST and k- ε realisable models are minimal and get less as the rotational speed rises, whereas the k-ε realisable model predicts more. Nonetheless, the simulation utilized frequencies of 3 Hz, 6 Hz, and 10 Hz for the engine, these are somewhat lower than the actual processes in the Stirling cycle. Furthermore, the CFD models for the cooler and heater assumed isothermal behaviour.

Following Caetano et al.'s [41] suggestion, a steady-state CFD simulation and Schmidt's model were integrated to generate the initial conditions and boundary conditions for investigation of Stirling engine of this type. The authors also emphasises that while CFD models are more accurate and reliable, they nevertheless require input data that is challenging to gather experimentally. The technique not only resulted in a sizable reduction in computing time but also improved the precision of the suggested power forecast. The author emphasizes the use of a CFD model, acknowledged for its reliability despite requiring significant computer resources. Furthermore, the lack of a well-defined methodology in the literature for determining the displacer's temperature profile and extrapolating initial conditions is mentioned.

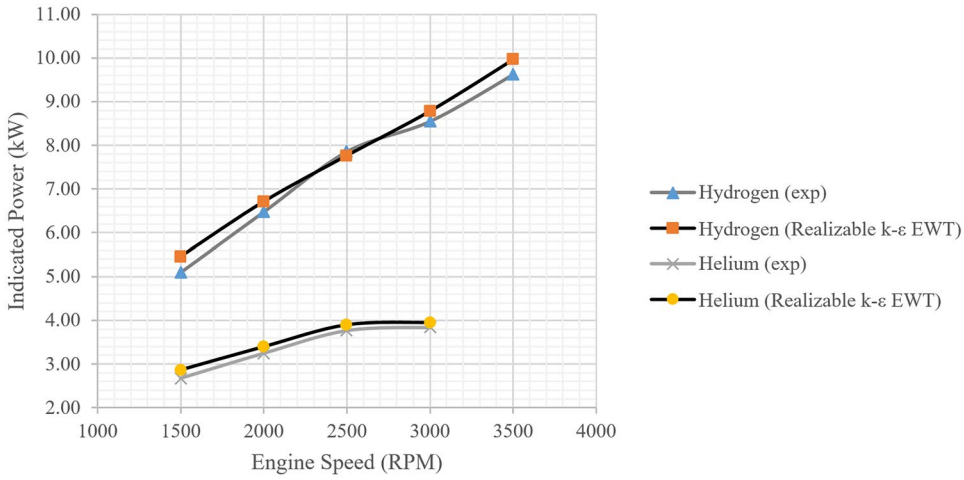


Fig. 1 Indicated power Vs engine speed for simulation and experimental results for hydrogen and helium as working fluid [42]

With a detailed geometry of the real engine, El-Ghafour et al. [42] conducted a 3-dimension Simulation in CFD on the GPU-3 Stirling engine. Six turbulent models were used to simulate the engine and compared against the experimental data, realizable k-ε model with enhanced wall treatment(EWT) was found to provide the most accurate result, around a 4% difference. The tumble vortices's significant impinging effect which is generated by the flow jetting results in primary heat transfer rates occurring within the compression and expansion regions during the expansion strokes. Regrettably, the investigation neglected the complex and conjugate heat transfer occurring within the interconnected fluid-solid domain in the porous media regenerator and the cold side heat exchanger. This oversight might have introduced inaccuracies in predicting heat transfer. The initial temperature parameters for the regenerator, cooler, and heater were set without taking this crucial aspect into account. Few results of this study are represented in Fig. 1.

This figure indicates the indicated power of GPU-3 engine at various engine speeds. The performance increases with increase in engine speed. The working fluid affects the performance of engine, hydrogen has better performance than helium because of lower molecular weight, higher thermal conductivity and lower specific heat capacity. The experimental vs. simulated data is also represented, the model used is realizable k-ε EWT which is in close agreement with the experimental data.

Rogdakis et al. [43] validated a 3 dimensioal CFD model for a Stirling engine of β-type using a confirmed 1D third-order model. The simulated engine used in the study is similar to GPU-3 Stirling engine but not identical to it. The model predicted a power output of 6.93 kW power and an efficiency of 37%.

Cheng and Phung [44] performed a CFD parametric study on 100 W Stirling engine and validated it with an experiment. In contrast to the previous conventional k-ε model, they adjusted the speed of rotation, charge pressure, heating and cooling temperatures, and length of piston. The simulation findings demonstrate that the indicated power (or charged pressure) is directly proportionate to the total mass. Furthermore, while the charged pressure is relatively low, there is an increase in thermal efficiency. As the charge pressure surpasses its critical value, thermal efficiency experiences a decline due to the accelerated increase in the rate of input of heat compared to the indicated power in the high-charged pressure domain.

A Free-Piston Gamma Stirling engine was subjected to multidimensional CFD modelling by Zhao et al. [45], which was compared with numerical modelling carried out using Sage software. The cold heat exchanger exhibited a pressure drop variance of 10.43%, and the temperature profile of the regenerator's solid matrix displayed an almost linear trend in the axial flow direction. The majority of the engine's components were created via additive manufacturing. The advanced free-piston engine's oscillating flow and heat transfer are presumed to be compressible, non-isothermal, transient, and turbulent in this study.

Cheng and Phung [46] proposed a novel method for exchanging data between a 2nd order thermodynamic model and a 3D CFD model, which drastically reduced the simulation time. The study was conducted on a 100 W beta Stirling engine with an external regenerator and also validated against the experimental data. A realizable k-ε turbulence model is used. It was discovered that the regenerator takes a lot longer than other chambers to reach the temperature of the solid matrix and the working fluid. After starting CFD with an initial estimate, data is recorded and transferred to a thermodynamic model after a few cycles, and when the discrepancy between the two models' stated power and thermal efficiency is less than 3%, subchamber temperatures and pressure are used to update the CFD model. This advances the model into a stable regime quickly. Cost in CFD is significantly reduced from 227 h to 80 h by this method. 3D

CFD studies performed on Stirling engines are summarised in Table 2.

Regenerator is a critical component of a Stirling engine, it has an impact on temperature distribution and heat transfer. The Local Thermal Non-Equilibrium model (LTNE) is important for the accurate modelling of porous regenerators and the radiation model is important for the accuracy of studies of the medium to high-temperature Stirling engine. However Ansys Fluent has a limitation regarding the simultaneous use of the Porous media Local Non-Equilibrium Model (LTNE) with the Radiation model. Standard radiation models like Discrete Ordinates and Surface to Surface cannot be used with LTNE whereas models like the Discrete Transfer Radiation Model and Monte Carlo Ray Tracing Method can be used with LTNE [47]. This will impact the accuracy of simulation studies involving porous regenerators and studies involving solar energy.

3 Equations

Table 3 presents deviation, between simulated and experimental Indicated Power, in different eddy viscosity models, which are used in this study. It also lists the domain of the study with reference. A CFD study depends on various parameters, one of the important is the turbulence model used, it is evident from the table that standard k-ε and realizable k-ε models have good accuracy. Thus these models can be effective for designing and evaluating performance of the engine.

Equations used in these models are [42, 48]:

The Unsteady Reynolds-Averaged Navier-Stokes (URANS) equations, in tensor form, can be obtained by employing a Cartesian coordinate system and breaking down the flow variables into ensemble-averaged values u, T, and p, and their corresponding fluctuations, u', T', p'.

Equation for conservation of mass:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_i}\left(\rho u_{ri}\right) = 0 \tag{1}$$

Equations for conservation of momentum:

$$\frac{\partial\left(\rho \cdot u_i\right)}{\partial t} + \frac{\partial\left(\rho \cdot u_{rj} \cdot u_i\right)}{\partial x_j} = -\frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j}\left[\mu \cdot \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) + \lambda \cdot \delta_{ij} \cdot \frac{\partial u_k}{\partial x_k} - \rho \cdot \overline{u_i' \cdot u_j'}\right] \tag{2}$$

where λ is the bulk viscosity coefficient and $_{ij}$ is Kronecker-Delta operator, taking the value of one if i = j and zero if i ≠ j.

where k is turbulent kinetic energy.

Equation for conservation of energy:

$$\frac{\partial(\rho \cdot E)}{\partial t} + \frac{\partial}{\partial x_i}\left[u_{ri}(\rho \cdot E + p)\right] = \frac{\partial}{\partial x_i}\left[K_{eff} \cdot \frac{\partial T}{\partial x_i} - \rho \cdot C_p \overline{u_i \cdot T'}\right] + \varphi \tag{3}$$

where the viscous dissipation term is while the effective thermal conductivity is K_{eff}, which is defined as:

$$K_{eff} = K + K_t \tag{4}$$

where K and K_t are the laminar and turbulent thermal conductivity, respectively.

Standard k-ε Model

$$\frac{\partial}{\partial t}(\rho k) + \frac{\partial}{\partial x_i}\left(\rho k u_i\right) = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_k}\right)\frac{\partial k}{\partial x_j}\right] + G_k + G_b - \rho\varepsilon - Y_M + S_k \tag{5}$$

and

$$\frac{\partial}{\partial t}(\rho \varepsilon) + \frac{\partial}{\partial x_i}\left(\rho \varepsilon u_i\right) = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial x_j}\right] + C_{1\varepsilon}\frac{\varepsilon}{k}(G_k + C_{3\varepsilon}G_b) - C_{2\varepsilon}\rho\frac{\varepsilon^2}{k} + S_\varepsilon \tag{6}$$

where, G_k stands for turbulence kinetic energy generated because of the mean velocity gradients, G_b stands for the turbulence kinetic energy generated because of buoyancy. The fluctuating dilatation in compressible turbulence to the overall dissipation rate is contributed by Y_M. The turbulent Prandtl number for k is σ_k and for ε is σ_ε. $C_{1\varepsilon}$, $C_{2\varepsilon}$, $C_{3\varepsilon}$ are constants S_k and S_ε are source terms defined by the user.

Realizable k-ε with Enhanced Wall Treatment (EWT)

$$\frac{\partial}{\partial t}(\rho k) + \frac{\partial}{\partial x_j}\left(\rho k u_j\right) = \frac{\partial}{\partial x_j}\left(\left(\mu + \frac{\mu_t}{\sigma_k}\right)\frac{\partial k}{\partial x_j}\right) + G_k + G_b - \rho\varepsilon - Y_M + S_k \tag{7}$$

and

$$\frac{\partial}{\partial t}(\rho\varepsilon) + \frac{\partial}{\partial x_j}\left(\rho\varepsilon u_j\right) = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial x_j}\right] + \rho C_1 S\varepsilon - \rho C_2 \frac{\varepsilon^2}{k + \sqrt{\nu\varepsilon}} + C_{1\varepsilon}\frac{\varepsilon}{k}C_{3\varepsilon}G_b + S_\varepsilon \tag{8}$$

where

$$C_1 = max\left[0.43, \frac{\eta}{\eta + 5}\right],\ \eta = S\frac{k}{\varepsilon},\ S = \sqrt{2S_{ij}S_{ij}}$$

where, G_k stands for turbulence kinetic energy generated because of the mean velocity gradients, G_b stands for the turbulence kinetic energy generated because of buoyancy. The fluctuating dilatation in compressible turbulence to the overall dissipation rate is contributed by Y_M. The turbulent Prandtl number for k is σ_k and for ε is σ_ε. $C_{1\varepsilon}$ and $C_2$are constants S_k and S_ε are source terms defined by the user.

Renormalization group (RNG) k-ε EWT

$$\frac{\partial}{\partial t}(\rho k) + \frac{\partial}{\partial x_i}\left(\rho k u_i\right) = \frac{\partial}{\partial x_j}\left(\alpha_k \mu_{eff}\frac{\partial k}{\partial x_j}\right) + G_k + G_b - \rho\varepsilon - Y_M + S_k \tag{9}$$

Table 2 Summary of 3D CFD studies performed on Stirling Engines

Author	Year	Engine Type	Model used	Domain	Software used	Remarks
Chen et al. [34]	2014	Gamma	Laminar	Quarter-scaled down to half	Custom	• Studied a Low-temperature difference engine, which had no external regenerator. • Power of engine is 0.62 W and efficiency is 4.82%.
Almajri et al. [35]	2017	Alpha	Laminar	Full	COMSOL	• Performed 3D CFD parametric study. • Power output increased by 65%, from 146 to 240 W
Mahkamov [36]	2006	Gamma	k-ε Turbulence	Full	Ansys Fluent	• Study of a biomass-powered Stirling engine. • Deviation for pressure and temperature is between 12–18%.
Li et al. [37]	2014	Alpha	Laminar	Full	Ansys Fluent	• Evaluated performance for different hot heat exchangers and regenerator geometries. • Regular-shaped microchannels are better than wire mesh regenerators.
Cheng, Chen [38]	2017	Beta	Standard k-ε Turbulence	Full	Ansys Fluent	• Trend of numerical predictions agrees well with experimental data. • Actual values differ by over 27%.
Xiao et al. [39]	2017	Beta	Standard k-ε Turbulence	Simplified	Ansys Fluent	• A maximum deviation of 13.4% with experimental data. • Optimised design had lower dead space with low flow resistance loss, increasing the power output by 40%.
Kuban et al. [40]	2018	Alpha	Realizable k-ε Turbulence	Half	Ansys Fluent	• In comparison to the standard version, the realizable form of the k-ε turbulence model exhibits superior capability in capturing intricate flow structures that encompass separation, rotation, and even recirculation regions. • Helium is better than Nitrogen and Nitrogen is better than air as the working fluid.
Caetano et al. [41],	2019	Beta	k-ω SST	Half	Ansys Fluent	• Proposed a combination of Schmidt's model and CFD simulation for quicker simulation of the Engine.
El-Ghafour et al. [42]	2019	Beta	Realizable k-ε EWT	One eighth	Ansys Fluent	• Compare 6 different eddy viscosity models for simulation. • Realizable k- ε model with Enhanced wall treatment (EWT) provides the most accurate results. • Average deviation is 4% compared to experimental data.
Rogdakis et al. [43]	2019	Beta	Standard k-ε Turbulence	One eighth	Ansys Fluent	• Compared validated 1D third order method with CFD method. • Deviation of about 2% in Engine efficiency compared to experimental data.

Table 2 (continued)

Author	Year	Engine Type	Model used	Domain	Software used	Remarks
Cheng et al. [44]	2020	Beta	Realizable k-ε Turbulence	Full	Ansys Fluent	• Maximum thermal efficiency of the engine was 41.9% at 10 bar charge pressure.
Zhao et al. [45]	2021	Free piston Gamma	Realizable k-ε Turbulence	One fifteenth	Ansys Fluent	• Compared CFD modelling of Free Piston Gamma Stirling with, numerical modelling done on Sage Software. • Deviation of 10.43% was noted between them. • Direct Numerical Simulation (DNS) and Large Eddy Simulation (LES) can provide more accurate results than Reynolds-Averaged Navier-Stokes (RANS) method. • But it is difficult to use them in complex engineering problems therefore RANS is used.
Cheng et al. [46]	2021	Beta	Realizable k-ε Turbulence	Full	Ansys Fluent	• Proposed a novel method to exchange data between the second-order thermodynamic model and 3D CFD model, to reduce the simulation time. • CFD simulation time reduced from 227 h to 80 h (~ 184% reduction).

Table 3 Deviation in indicated power in various eddy viscosity models

Model	Deviation in Power (simulated Vs experimental)	Domain	Reference
Laminar	9%	2D planar	[30]
	7%, 18% (w/o radiation)	2D axis-symmetric	[16]
Standard k-ε	8.1%	3D one eighth	[42]
	12%	3D full	[36]
	13.7%	3D simplified	[39]
	27%	3D full	[38]
	40%	2D axis-symmetric	[28]
	3.8%	3D one eighth	[42]
Realizable k-ε EWT	4.4%	3D one eighth	[42]
RNG k-ε EWT	15.1%	2D axis-symmetric	[32]
SST k-ω	6.3%	3D one eighth	[42]
SST k-ω low Re correct			

and

$$\frac{\partial}{\partial t}(\rho\varepsilon)+\frac{\partial}{\partial x_i}\left(\rho\varepsilon u_i\right)=\frac{\partial}{\partial x_j}\left[\alpha_\varepsilon\mu_{eff}\frac{\partial\varepsilon}{\partial x_j}\right]+C_{1\varepsilon}\frac{\varepsilon}{k}(G_k+C_{3\varepsilon}G_b)-C_{2\varepsilon}\rho\frac{\varepsilon^2}{k}-R_\varepsilon+S_\varepsilon \tag{10}$$

where, G_k stands for turbulence kinetic energy generated because of the mean velocity gradients, G_b is the generation of turbulence kinetic energy due to buoyancy. Y_M represents the contribution of the fluctuating dilatation in compressible turbulence to the overall dissipation rate. The inverse effective Prandtl numbers for k and ε, are the quantities α_k and α_ε are respectively. S_k and S_ε are user-defined source terms.

Shear-Stress Transport (SST) k-ω

$$\frac{\partial}{\partial t}(\rho k)+\frac{\partial}{\partial x_i}\left(\rho k u_i\right)=\frac{\partial}{\partial x_j}\left[\Gamma_k\frac{\partial k}{\partial x_j}\right]+G_k-Y_k+S_k \tag{11}$$

and

$$\frac{\partial}{\partial t}(\rho\omega) + \frac{\partial}{\partial x_j}\left(\rho\omega u_j\right) = \frac{\partial}{\partial x_j}\left[\Gamma_\omega \frac{\partial \omega}{\partial x_j}\right] + G_\omega - Y_\omega + D_\omega + S_\omega \tag{12}$$

where, production of turbulence kinetic energy is represented by G_k, the generation of ω is represented by G_ω. Γ_k represents the effective diffusivity of k and Γ_ω represents the effective diffusivity of ω. D_ω represents the cross-diffusion term. Y_k and Y_ω represent the dissipation of k and ω due to turbulence. S_k and S_ε are user-defined source terms [47].

4 Conclusions

The selection of numerical schemes, discretization and integration methods, time step size, and computational mesh quality significantly influence the sensitivity of CFD analysis. Therefore, careful and appropriate choices are crucial for achieving stable results and convergence. A few more notable points about CFD study in Stirling engine are:-.

- CFD approach stands as the sole method capable of modeling conjugated flow and heat transfer processes while taking into account real geometries.
- The flow within nearly all engine components exhibits spatially strong non-uniformity, with pressure and temperature distributions being inherently inhomogeneous. Therefore, Full domain 3D CFD study is preferred over partial domain 3D and 2D CFD studies.
- Iimpinging effect is noticed in beta-type and gamma-type Stirling engines which leads to the formation of vortices and enhances heat transfer. These vortices can be predicted only by CFD analysis. This is evident in the discrepancy, of over 80% [26], in power output predicted by the second order and CFD method.
- Accuracy of the CFD study is improved by 11% by considering internal radiation heat transfer.
- Most of the studies neglect mechanical friction and internal radiation heat transfer.
- ANSYS Fluent is the most popular choice among researchers for CFD analysis of the Stirling engine.
- ANSYS Fluent is limited regarding the simultaneous use of the Porous media Local Non-Equilibrium Model (LTNE) with the Radiation model. Standard radiation models like Discrete Ordinates and Surface to Surface cannot used with LTNE whereas models like Discrete Transfer Radiation Model and Monte Carlo Ray Tracing Method can used with LTNE. This will impact the accuracy of simulation studies involving porous regenerators and studies involving solar energy.

Stirling engines can be vital in the future energy generation in hybrid as well as renewable systems, such systems are sensitive to the optimal size of the engine, which is very sensitive to design changes. Performing these optimisations while incorporating CFD will drastically improve the design of the engine. One such example is of a Solar Dish Stirling Engine, which is sensitive to the location of the engine having a direct impact on Solar Irradiation availability. Performing CFD analysis is computationally intensive but improves accuracy, more research on methods that reduce the computational demand while retaining the accuracy is recommended.

Authors' Contributions AK: Conceptualization, and supervision, Writing—review & editing, and supervision. VS: Visualization methodology, Writing—original draft and investigation.

Funding There is no funding.

Data Availability All data are given in either Table or figure.

Code Availability (software application or custom code): Not Applicable.

Declarations

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

Ethical Approval (include appropriate approvals or waivers): Not Applicable.

Consent to Participate (include appropriate statements): Not Applicable.

Consent for publication (include appropriate statements): Not Applicable.

References

1. Frankel JA (2010) The Natural Resource Curse: A Survey. https://doi.org/10.3386/W15836
2. Höök M, Tang X (2013) Depletion of fossil fuels and anthropogenic climate change—A review. Energy Policy 52:797–809. https://doi.org/10.1016/J.ENPOL.2012.10.046
3. Bhandari B, Lee KT, Lee CS et al (2014) A novel off-grid hybrid power system comprised of solar photovoltaic, wind, and hydro energy sources. Appl Energy 133:236–242. https://doi.org/10.1016/J.APENERGY.2014.07.033
4. Singh UR, Kumar A (2018) Review on solar Stirling engine: development and performance. Therm Sci Eng Progress 8:244–256. https://doi.org/10.1016/j.tsep.2018.08.016
5. Sandoval OR, Caetano BC, Borges MU et al (2019) Modelling, simulation and thermal analysis of a solar dish/Stirling system: a case study in Natal, Brazil. Energy Convers Manag 181:189–201. https://doi.org/10.1016/J.ENCONMAN.2018.12.005
6. Zayed ME, Zhao J, Elsheikh AH et al (2020) Optimal design parameters and performance optimization of thermodynamically balanced dish/Stirling concentrated solar power system

using multi-objective particle swarm optimization. Appl Therm Eng 178:115539. https://doi.org/10.1016/j.applthermaleng.2020.115539
7. Zayed ME, Zhao J, Elsheikh AH et al (2021) A comprehensive review on Dish/Stirling concentrated solar power systems: design, optical and geometrical analyses, thermal performance assessment, and applications. J Clean Prod 283:124664. https://doi.org/10.1016/j.jclepro.2020.124664
8. Bataineh K (2022) Performance evaluation of a stand-alone solar dish Stirling system for off–grid electrification. Energy Sour Part A Recover Utilization Environ Eff 44:1208–1226. https://doi.org/10.1080/15567036.2022.2053249
9. Guarino S, Buscemi A, Messineo A, Lo Brano V (2022) Energy and Environmental Assessment of a Hybrid Dish-Stirling concentrating Solar Power Plant. Sustainability 14:6098. https://doi.org/10.3390/su14106098
10. Shboul B, AL-Arfi I, Michailos S et al (2021) Design and techno-economic assessment of a new hybrid system of a solar dish Stirling engine instegrated with a horizontal axis wind turbine for microgrid power generation. Energy Convers Manag 245:114587. https://doi.org/10.1016/j.enconman.2021.114587
11. Bekele EA, Ancha VR (2022) Transient performance prediction of solar dish concentrator integrated with stirling and TEG for small scale irrigation system: a case of Ethiopia. Heliyon 8:e10629. https://doi.org/10.1016/j.heliyon.2022.e10629
12. Geng D, Cui J, Fan L (2021) Performance investigation of a reverse osmosis desalination system powered by solar dish-Stirling engine. Energy Rep 7:3844–3856. https://doi.org/10.1016/j.egyr.2021.06.072
13. Walker G (1973) The Stirling Engine. Sci Am 229:80–87. https://doi.org/10.1038/scientificamerican0873-80
14. Walker G, Senft JR (1985) Free-Piston Stirling Engines. 23–99. https://doi.org/10.1007/978-3-642-82526-2_2
15. Walker G (1979) ELEMENTARY DESIGN GUIDELINES FOR STIRLING ENGINES. Proc 14th Intersoc Energy Convers Eng Conf 1:1066–1068
16. Ben-Mansour R, Abuelyamen A, Mokheimer EMA (2017) CFD analysis of radiation impact on Stirling engine performance. Energy Convers Manag 152:354–365. https://doi.org/10.1016/j.enconman.2017.09.056
17. Beale WT (1969) Free Piston Stirling engines - some model tests and simulations. SAE Tech Papers. https://doi.org/10.4271/690230
18. IWAMOTO S, HIRATA K, TODA F (2001) Performance of Stirling Engines. Arranging method of experimental results and performance prediction. JSME Int J Ser B 44:140–147. https://doi.org/10.1299/jsmeb.44.140
19. Babaelahi M, Sayyaadi H (2015) A new thermal model based on polytropic numerical simulation of Stirling engines. Appl Energy 141:143–159. https://doi.org/10.1016/J.APENERGY.2014.12.033
20. Bataineh KM, Maqableh MF (2022) A new numerical thermodynamic model for a beta-type Stirling engine with a rhombic drive. Therm Sci Eng Progress 28:101071. https://doi.org/10.1016/J.TSEP.2021.101071
21. Yang C, Zhuang N, Du W et al (2022) Modified Stirling cycle thermodynamic model IPD-MSM and its application. Energy Convers Manag 260:115630. https://doi.org/10.1016/J.ENCONMAN.2022.115630
22. Qiu H, Wang K, Yu P et al (2021) A third-order numerical model and transient characterization of a β-type Stirling engine. Energy 222:119973. https://doi.org/10.1016/J.ENERGY.2021.119973
23. Urieli I, Rallis CJ, Berchowitz DM et al (1977) Computer simulation of Stirling cycle machines. In: 12th IECEC. p Paper No. 779252
24. Aksoy F, Solmaz H, Karabulut H et al (2016) A thermodynamic approach to compare the performance of rhombic-drive and crank-drive mechanisms for a beta-type Stirling engine. Appl Therm Eng 93:359–367. https://doi.org/10.1016/J.APPLTHERMALENG.2015.09.105
25. Ansys Fluent | Fluid Simulation Software. https://www.ansys.com/en-in/products/fluids/ansys-fluent. Accessed 27 Apr 2023
26. COMSOL - Software for Multiphysics Simulation. https://www.comsol.com/. Accessed 6 Nov 2023
27. OpenFOAM. https://www.openfoam.com/. Accessed 6 Nov 2023
28. Mahkamov K (2006) An Axisymmetric Computational Fluid Dynamics Approach to the analysis of the Working process of a Solar Stirling Engine. J Sol Energy Eng 128:45–53. https://doi.org/10.1115/1.2148979
29. Salazar JL, Chen W-L (2014) A computational fluid dynamics study on the heat transfer characteristics of the working cycle of a β -type Stirling engine. Energy Convers Manag 88:177–188. https://doi.org/10.1016/j.enconman.2014.08.040
30. Alfarawi S, AL-Dadah R, Mahmoud S (2016) Influence of phase angle and dead volume on gamma-type Stirling engine power using CFD simulation. Energy Convers Manag 124:130–140. https://doi.org/10.1016/j.enconman.2016.07.016
31. Della Torre A, Guzzetti A, Montenegro G et al (2014) CFD modelling of a beta-type stirling machine. In: 11th World Congress on Computational Mechanics, WCCM 2014, 5th European Conference on Computational Mechanics, ECCM 2014 and 6th European Conference on Computational Fluid Dynamics, ECFD 2014
32. Mohammadi MA, Jafarian A (2018) CFD simulation to investigate hydrodynamics of oscillating flow in a beta-type Stirling engine. Energy 153:287–300. https://doi.org/10.1016/j.energy.2018.04.017
33. Abuelyamen A, Ben-Mansour R (2018) Energy efficiency comparison of Stirling engine types (α, β, and γ) using detailed CFD modeling. Int J Therm Sci 132:411–423. https://doi.org/10.1016/j.ijthermalsci.2018.06.026
34. Chen W-L, Wong K-L, Chang Y-F (2014) A computational fluid dynamics study on the heat transfer characteristics of the working cycle of a low-temperature-differential γ-type Stirling engine. Int J Heat Mass Transf 75:145–155. https://doi.org/10.1016/j.ijheatmasstransfer.2014.03.055
35. Almajri AK, Mahmoud S, Al-Dadah R (2017) Modelling and parametric study of an efficient alpha type Stirling engine performance based on 3D CFD analysis. Energy Convers Manag 145:93–106. https://doi.org/10.1016/j.enconman.2017.04.073
36. Mahkamov K (2006) Design improvements to a Biomass Stirling Engine using Mathematical Analysis and 3D CFD modeling. J Energy Resour Technol 128:203–215. https://doi.org/10.1115/1.2213273
37. Li Z, Haramura Y, Kato Y, Tang D (2014) Analysis of a high performance model Stirling engine with compact porous-sheets heat exchangers. Energy 64:31–43. https://doi.org/10.1016/j.energy.2013.11.041
38. Cheng C-H, Chen Y-F (2017) Numerical simulation of thermal and flow fields inside a 1-kW beta-type Stirling engine. Appl Therm Eng 121:554–561. https://doi.org/10.1016/j.applthermaleng.2017.04.105
39. Xiao G, Sultan U, Ni M et al (2017) Design optimization with computational fluid dynamic analysis of β-type Stirling engine. Appl Therm Eng 113:87–102. https://doi.org/10.1016/j.applthermaleng.2016.10.063
40. Kuban L, Stempka J, Tyliszczak A (2019) A 3D-CFD study of a γ-type Stirling engine. Energy 169:142–159. https://doi.org/10.1016/j.energy.2018.12.009

41. Caetano BC, Lara IF, Borges MU et al (2019) A novel methodology on beta-type Stirling engine simulation using CFD. Energy Convers Manag 184:510–520. https://doi.org/10.1016/j.enconman.2019.01.075
42. El-Ghafour SA, El-Ghandour M, Mikhael NN (2019) Three-dimensional computational fluid dynamics simulation of stirling engine. Energy Convers Manag 180:533–549. https://doi.org/10.1016/j.enconman.2018.10.103
43. Rogdakis E, Bitsikas P, Dogkas G, Antonakos G (2019) Three-dimensional CFD study of a β-type Stirling Engine. Therm Sci Eng Progress 11:302–316. https://doi.org/10.1016/j.tsep.2019.04.012
44. Cheng C, Phung D (2021) Numerical and experimental study of a compact < scp > 100-W-class β-type Stirling engine. Int J Energy Res 45:6784–6799. https://doi.org/10.1002/er.6271
45. Zhao W, Li R, Li H et al (2021) Numerical analysis of fluid dynamics and thermodynamics in a stirling engine. Appl Therm Eng 189:116727. https://doi.org/10.1016/j.applthermaleng.2021.116727
46. Cheng C, Phung D (2021) Exchanging data between computational fluid dynamic and thermodynamic models for improving numerical analysis of Stirling engines. Energy Sci Eng 9:2177–2190. https://doi.org/10.1002/ese3.965
47. (2022) ANSYS FLUENT user's guide, 2022 R1. ANSYS Inc.
48. (2016) ANSYS FLUENT Theory Guide., Release 17. ANSYS Inc.

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.

A Systematic Literature Review on Maintenance of Software Containers

RUCHIKA MALHOTRA

Delhi Technological University, Delhi, India

ANJALI BANSAL

Delhi Technological University, Delhi, India

MAROUANE KESSENTINI

Oakland University, Michigan, United States

Nowadays, cloud computing is gaining tremendous attention to deliver information via the internet. Virtualization plays a major role in cloud computing as it deploys multiple virtual machines on the same physical machine and thus results in improving resource utilization. Hypervisor-based virtualization and containerization are two commonly used approaches in operating system virtualization. In this paper, we provide a systematic literature review on various phases in maintenance of containers that are container image detection, container scheduling, container security measures, and performance evaluation of containers. We have selected 145 primary studies out of which 24% of studies are related to container performance evaluation, 42% of studies are related to container scheduling techniques, 22% of studies are related to container security measures, and 12% of studies are related to container image detection process. A few studies are related to container image detection process and evaluation of container security measures. Resource utilization is the most considered performance objectives in almost all container scheduling techniques. We conclude that there is a need to introduce new tagging approaches, smell detection approaches, and also new approaches to detect and resolve threat issues in containers so that we can maintain the security of containers.

Additional Keywords and Phrases: Containers, Docker, Orchestration, Performance evaluation, Scheduling techniques, Vulnerability, Kubernetes.

1 INTRODUCTION

The area of cloud computing is gaining tremendous attention nowadays. It manages the resources dynamically and automatically. Without virtualization, there is no existence of cloud computing as it reduces the hardware maintenance overhead [36]. We can achieve virtualization with the help of virtual machines and containers. Containers are replacing virtual machines as containers provide shared host operating system, fast deployment, quick launch time, portability, and scalability. A container is a closed environment for the software that bundles all binaries, library files, and supported files needed by the application in one package to provide a platform-independent run time environment that improves portability and productivity [6]. Containers are lightweight virtualization technologies. Containerization is the process of executing multiple applications on the same operating system or machine. In hypervisor-based virtualization, each application requires a separate operating system but in containerization, applications share an operating system, so the container-based virtualization technique is light weighted and deployment can be smaller in size. The main benefits of container-based virtualization are it virtualizes operating system, takes a few seconds to run, is smaller in size, is lightweight, and requires less memory. The main drawback of containerization is containers are less secure than virtual machines. As containers use host operating system so if we restart the container then there is no need to reboot or restart the host operating system.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM 0360-0300/2024/02-ART
http://dx.doi.org/10.1145/3645092

Just Accepted

1.1 Containerization Technologies

In this section, we describe some container technologies such as docker, LXC, OpenVZ, singularity, and uDocker that represents container-based virtualization techniques. Now a days, these technologies are used by scientific computing communities. Container technology such as Docker focuses on specific industry applicability and singularity focuses on container portability across high performance computing environments. These technologies execute its own set of methods to achieve network isolation. Table 1 shows the comparative study of various containerization technologies such as docker, lxc, openvz, singularity, and udocker.

Table 1. Containerization Technologies

Docker	**LXC (Linux Container)**	**Open VZ (Open Virtuozzo)**	**Singularity**	**uDocker**
Introduced by a company named as Docker, Inc.	Introduced by Daniel Lezcano and Serge Hallyn.	Introduced a company named as SWsoft which is currently named as Parallels.	Introduced by Gregory M. Kurtzer.	Introduced by a team at CERN (European Organization for Nuclear Research) team.
Enhanced version of LXC. Docker containers share host OS but runs in isolated user space. They are used in DevOps and application deployment scenarios.	LXC directly operates on linux kernel and is used for running multiple containers on a single host.	It has its own kernel (virtuozzo kernel) that it uses for containerization. Same as LXC, OpenVZ also runs multiple containers on a single host.	Singularity provides the facility to run the containers without requiring root access and used on high performance computing clusters and supercomputers.	uDocker is used where users need to execute Docker containers without requiring root access or other administrative rights.
It is used for application-level containerization.	It is used for system level containerization.	Same as LXC. It is used for system level containerization.	It is used for application-level containerization	It is used for application-level containerization.
Faster deployment speed	Deployment speed is slower as compared to docker and singularity	Deployment speed is slower as compared to docker and singularity	Deployment speed is slower as it takes longer startup time than docker	Deployment speed is slower than others.
Supported on all platforms such as windows, linux, macOS, etc.	Supported on linux platform only.	Supported on linux platform only.	Supported on linux platform only.	Supported on linux platforms only.

Just Accepted

1.2 Container Orchestration Tools

In this section, we describe container orchestration tools. These tools are used to manage application deployment, load balancing, allocation of resources, scheduling, maintaining secure communication between containers, and scaling [15][4]. Container-based virtualization uses many orchestration tools such as Docker swarm, Kubernetes, Mesos, Nextflow, Cloudify, Openshift, and Rancher. A workflow management system called Nextflow was created for data-driven computational pipelines. Nextflow is intended to provide a flexible, reproducible, and portable workflow management solution running on multiple platforms such as cloud platforms and high-performance computing. Table 2 describes the comparative study of docker swarm, Kubernetes, and mesos.

Table 2. Container Orchestration Tools

Docker Swarm	**Kubernetes**	**Mesos**
Docker swarm was introduced by Docker.	Kubernetes was introduced by Google	Mesos was introduced at University of California, Berkeley
Easy to integrate and setup, flexible API, but limited customization.	Highly versatile, large open-source dev community, but more expensive.	Best for large—systems and designed for maximum redundancy.
Running rate is limited to 1000 nodes	Running rate is limited to 1000 nodes	Can run upto 50000 nodes
Great value and easily scalable for small to medium systems.	Best for medium-scale, highly redundant systems, but requires a larger IT staff.	The most stable platform for large scale data solutions, but overly complex for small-scale systems under 10-20 nodes.
For simpler orchestration needs and smaller containerized applications, Docker Swarm is suitable. It is often selected when you want to incorporate clustered container management into your existing Docker deployment.	For complex microservices architectures and applications that require scalability, robustness, and fine-grained management of containerized workloads, Kubernetes is often used.	Mesos is appropriate for companies with a variety of workloads, including big data, containers, and other applications. It offers effective resource utilization for a variety of applications and resource isolation.

Container technology has been used in many fields such as Scientific computing, big data processing, high-performance computing, high throughput computing, cloud computing, fog computing, edge computing, the internet of things, DevOps, and grid computing.

Just Accepted

When there are only a few containers, container maintenance can be as simple as learning the build and run syntax for docker, signing up for a docker hub account, and launching your preferred text editor. However, if there are a large number of containers, it becomes necessary to maintain the software within the containers so that the IT team can easily manage them. The container maintenance strategy is broadly classified into four components: container image supply chain, container infrastructure and orchestration, container runtime security, and container observability. The container image supply chain defines how the containers are built, which files to include in the container, how to assign unique tags to the images, and tools that check for logical errors and smells in the files. The second step describes where and how the containers run together. Orchestration tools such as docker swarm, kubernetes, and mesos are used to schedule the containers on working nodes. The third step, container runtime security, deals with approaches to detect vulnerabilities in running containers, and how to handle vulnerabilities that are detected. The last step, container observability, is related to runtime metrics such as CPU usage, memory usage, network bandwidth usage, and disk usage. The existing literature surveys [6],[11],[12] focused on container scheduling techniques, various methods used to analyze the performance of containers, benchmark tools used, and container performance monitoring tools. This paper aims at providing a summative evaluation of existing literature studies on containerization technology to provide a better understanding of container security attacks, vulnerabilities, approaches to mitigate security vulnerabilities and attacks, various container scheduling techniques used in the literature, approaches to identify which image to be included in the container, image tagging approaches, benchmark tools used to evaluate performance, and various container performance monitoring tools. The rest of the paper is assembled as follows: Method for conducting a literature review (Research methodology) is described in section 2 which consists of defining research questions, inclusion and exclusion criteria, and quality assessment criteria. Section 3 describes the answer to the framed research questions. Section 4 describes the conclusion and future guidelines.

2 RESEARCH METHODOLOGY

This paper follows the guidelines suggested by [55],[80] which decomposes a systematic literature review into three phases: planning the review, conducting the review, and reporting the review results. We identify the need for systematic review and develop protocols for review in the first phase. In the second phase, we frame research questions that are needed to be addressed in the systematic literature review, quality assessment criteria to analyze the studies, inclusion and exclusion criteria, and study selection criteria that includes identification of relevant keywords and identifying the primary sources or databases to be search in order to identify the relevant studies. In the last phase, we report the review results.

In this section, we define the research questions underlying our survey, literature search strategy, quality assessment criteria, and inclusion and exclusion criteria.

2.1 Research Questions

Table 3. shows the research questions formed based on the objectives described in the introduction section. From the selected primary studies, we analyzed how many studies are related to container

Just Accepted

image detection, container scheduling techniques, container security measures, and container performance evaluation (RQ1). We analyzed studies related to container image detection to answer RQ 2, RQ 2.1, RQ 2.2, and RQ 2.3. RQ 3, RQ 3.1, RQ 3.2 and RQ 3.3 assess which techniques and which performance objectives have been used in scheduling the containers. In RQ 4, we analyzed container security vulnerabilities, container security attacks, and various approaches used in mitigating the vulnerabilities and attacks. The last question RQ 5 assess the studies related to performance evaluation of containers, benchmarks tools, and container performance monitoring tools.

Table 3. Research Questions

RQ 1	How many studies exist in the literature which deal with mining the container images, scheduling the containers, container security, and evaluating the performance of the containers?
RQ 2	Which container images, tagging approaches, smell detection approaches, and image mining approaches have been used in containerization?
RQ 2.1	What are the approaches used in container image tagging?
RQ 2.2	What are the various types of smell detected and the approaches used to detect the smell in container images?
RQ 2.3	What are the image mining approaches used in literature?
RQ 3	Which dataset, techniques, and performance objectives have been used in container scheduling?
RQ 3.1	Which type of dataset has been used in container scheduling?
RQ 3.2	What are the various performance objectives used in container scheduling?
RQ 3.3	Which technique has been used in container scheduling?
RQ 4	What are the various security vulnerabilities, vulnerability detection methods, attacks, and solutions to mitigate the vulnerability issues in containers?
RQ 4.1	What are the various security vulnerabilities in containers and which methods have been used to detect the vulnerabilities?
RQ 4.2	Which methods have been used to resolve the security vulnerability issues?
RQ 4.3	What are the various types of attacks based on attack layers (container, application, docker engine, and host) and approaches used to detect security attacks in containerized applications?
RQ 5	Which techniques, benchmarks, and monitoring tools have been used to evaluate the container performance?
RQ 5.1	What are the various performance evaluation techniques used to evaluate the performance of the containers?
RQ 5.2	Which benchmarks have been used in the performance evaluation of containers?

Just Accepted

RQ 5.2.1	What are the various CPU-intensive benchmarks?
RQ 5.2.2	What are the various memory-intensive benchmarks?
RQ 5.2.3	What are the various disk I/O intensive benchmarks?
RQ 5.2.4	What are the various network I/O intensive benchmarks?
RQ 5.3	Which tools have been used to monitor the container performance?

2.2 Search Strategy

We collect the primary studies from various scientific literature databases by using the search string "(Container OR Docker) AND (Scheduling OR Resource allocation OR Heuristic OR Metaheuristic OR Mathematical modeling OR Machine learning) AND Performance evaluation AND Monitoring tool AND Benchmark tool AND (Security OR Vulnerability OR Attacks) AND Image detection AND (Smells OR Antipatterns) AND Tagging". Here AND combines the main search terms and OR combines alternative terms and synonyms. After selecting the search terms, we identify the databases. We have used the following databases to search relevant studies [80]:

1. IEEE Xplore
2. Springer
3. ACM digital library
4. Google Scholar
5. Elsevier
6. Wiley
7. Taylor & Francis

After choosing which databases to search, a preliminary search was conducted to find the primary studies. After performing a preliminary search, the pertinent studies were found by obtaining the full-text papers following the inclusion and exclusion criteria outlined in the next section. We also included those studies that were found important from the reference section of the pertinent studies. The studies were chosen after following inclusion and exclusion criteria.

We identified 140 primary studies initially. After reviewing the reference list of the pertinent studies, further 40 studies were included. Hence, total number of 180 studies were identified for further analysis.

2.3 Inclusion and Exclusion criteria

The following inclusion and exclusion criteria is followed for including a study in this systematic literature review:

Inclusion criteria:

- The study must be related to computer science and engineering and propose techniques, methods, and tools for container-related problems.

Just Accepted

- The study must be written in English.
- The study must pass quality assessment criteria.
- If a conference paper has a significant journal extension, then we will include both conference and journal publications.

Exclusion criteria:

- Studies that are not related to computer science and engineering field.
- Thesis, poster presentation papers, demo paper, and short papers.

By applying the above inclusion and exclusion criteria, 160 studies were selected. The quality assessment criteria given in the section 2.4 was used to identify the final primary studies.

2.4 Quality Assessment Questions

We formed quality assessment questions for weighing the studies. These questions are ranked 1 or 0.5 or 0. The final score is calculated by adding the score assigned to each question. A study could score a maximum of 9 and a minimum of 0. Studies having a final score of 5 or more than 5 will be included in primary studies. After applying quality assessment criteria, we identified 145 primary studies. Table 4. shows quality assessment questionnaire.

Table 4. Quality Assessment Criteria

QA 1	Does the study state the main objectives and research questions in a clear way?
QA 2	Does the study clearly give the details about selected projects?
QA 3	Is the dataset clearly described (number of stars, contributors, services, and containers)?
QA 4	Does the study give a detailed description of selected techniques?
QA 5	Does the study clearly define the parameter settings of the techniques used?
QA 6	Are the performance objectives clearly stated?
QA 7	Does the study clearly report the results?
QA 8	Are the result compared with the results of previous studies?
QA 9	Does the study discuss about validity threats?

3 REVIEW RESULTS

In this section, we will answer all the defined research questions based on selected primary studies. The answers to the defined research questions are given in each sub sections.

The primary studies are categorized based on publication type and publication year as shown in figure 1. It shows that majority of the studies were published between 2017 and 2022. Out of 145 primary studies, 48 studies were published in top most journals such as IEEE Access, Future Generation Computer Systems, Cluster Computing, Soft Computing, etc., 87 studies were published in

Just Accepted

top conferences, and 10 studies were published as book chapters. Figure 2 depicts the distribution of primary studies by their publishers over the year between 2015 and 2022. It shows that majority of the studies have been published at IEEE, Springer, and ACM.

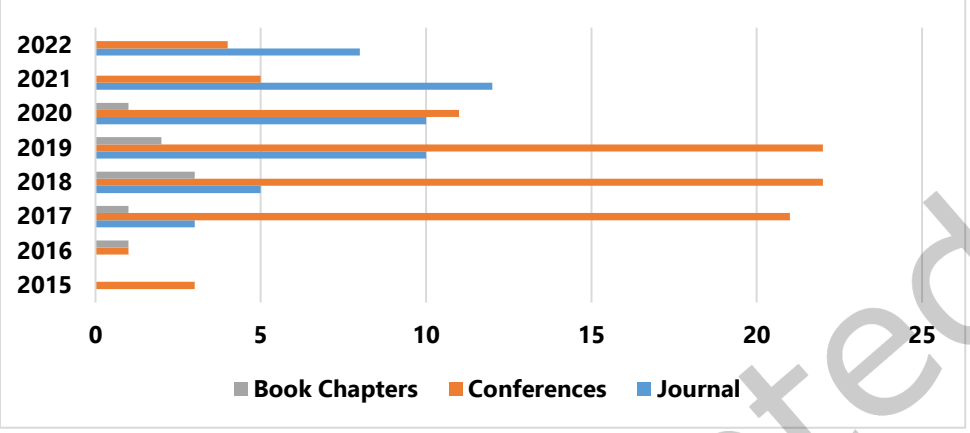


Fig. 1 Categorization based on Publication Type and Publication Year

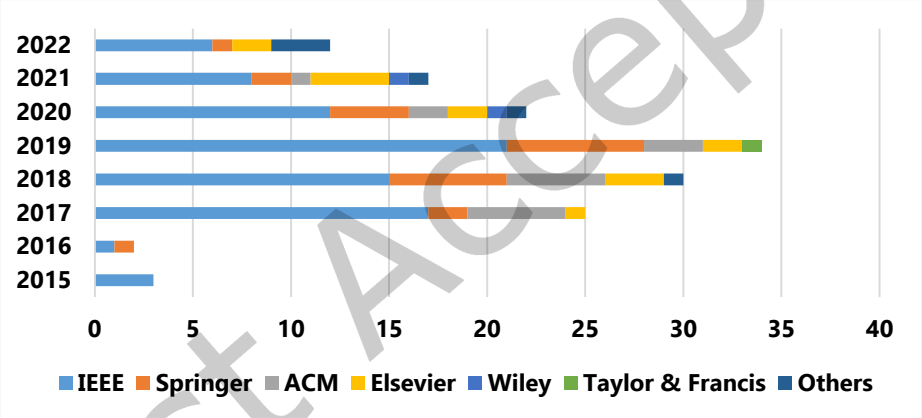


Fig. 2 Distribution of Primary Studies by Publisher Over Years

3.1 RQ 1 How many studies exist in the literature which deal with mining the container images, scheduling the containers, container security, and evaluating the performance of the containers?

Figure 3 describes pie chart of selected primary studies in each phase. This figure shows that out of 145 studies, 24% studies are related to performance evaluation of containers, 42% studies are related to container scheduling approaches, 22% studies are related to container security, and 12% studies are related to container image detection.

Just Accepted

Table 5. Classification of Studies

Image Detection	[21], [23], [26], [27], [37], [42], [65], [76], [108], [109], [126], [134], [135], [137], [139], [140], [150], [152].
Container Scheduling Techniques	[5],[7], [8], [9], [11], [22], [25], [28], [29], [30], [32], [33], [34], [38], [39], [40], [43], [49], [51], [52], [53], [54], [60], [61], [62], [63], [66], [69], [70], [73], [75], [77], [78], [82], [86], [88], [89], [90], [91], [92], [98], [99], [100], [102], [106], [110], [111], [113], [115], [117], [120], [121], [122], [123], [124], [125], [129], [133], [136], [145], [148], [149], [153].
Container Security	[10], [31], [35], [41], [44], [47], [50], [57], [59], [67], [68], [71], [74], [83], [93], [104], [116], [118], [119], [127], [128], [131], [132], [142], [143], [144], [146], [151].
Performance evaluation of containers	[1], [2], [3], [13], [14], [16], [17], [18], [19], [20], [24], [45], [46], [48], [56], [58], [64], [72], [79], [81], [84], [85], [87], [94], [95], [96], [97], [101], [103], [105], [107], [112], [114], [130], [138], [141], [147].

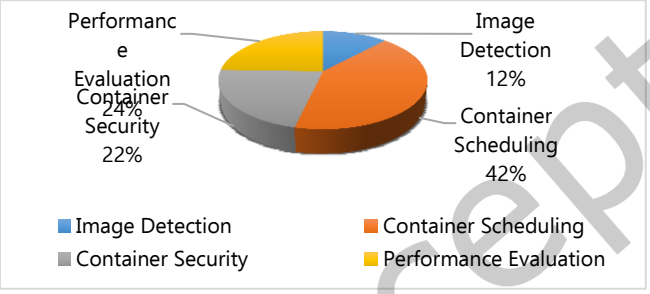


Fig. 3 Primary Studies Selected in Each Phase

Just Accepted

Discussion

- According to the analyzed studies, most papers were published between 2017 and 2022.
- Most of the studies were published in conference proceedings. Out of 145 studies, 89 studies were published as conference proceedings, 48 studies were published in journals, and 8 studies were published as book chapters.
- We have described 4 phases of software container maintenance: container image detection, container scheduling, container security, and container performance evaluation.
- Container scheduling is the most explored area in the analyzed papers. Out of 145 studies, 62 studies are related to container scheduling, 37 studies are related to container performance evaluation, 28 studies are related to container security, and 18 studies are related to container image detection.

3.2 RQ 2 Which container images, tagging approaches, smell detection approaches, and image mining approaches have been used in containerization?

In this section, we explain image mining approaches, smell detection approaches, and tagging approaches in containerizing an application. Lin et al. [65] conducted an empirical study in which they

explored base images, popular programming languages, some tagging practices, and various code smells and created a dataset by collecting data from github, docker hub, and bit bucket repositories. The authors used release tagging and SHA hash pinning approach and the result shows that the release tagging approach is more commonly used and the number of code smells in docker files is decreasing day by day which shows an improvement in the quality of docker images. Cito et al. [26] explored smells in Docker files and popular base images. This study was the first of its type on Docker files and highlighted various features of the Docker ecosystem and considered Docker files stored in the github repository. Ibrahim et al. [42] explored the differences between official and community docker images for 5 software systems and they considered 936 images for analysis which consist of 929 community images and 7 official images. The result of this study demonstrates that 49%, 26%, and 8% of the community Debian, Alpine, and Ubuntu images uses fewer resources than their respective official images and as compared to official images, 7% of the community images have minimum security flaws.

3.2.1 RQ 2.1 What are the approaches used in container image tagging?

There can be a large number of docker repositories that contain docker images and Docker files. So, there is a need to assign some tags to docker repositories so that necessary files can be easily retrieved. Manually assigning tags is time consuming and exhausting process. So, Yin et al. [140] proposed D-Tagger which is a tag recommendation approach and follows Labeled Latent Dirichlet Allocation Algorithm for recommending tags to docker repositories based on docker files. The result shows that D-Tagger achieves 0.675 accuracy in terms of Recall@5 and 0.712 accuracy in terms of Recall@10. Yin et al. [139] proposed STAR which stands for Specialized Tagging Approach for Docker Repository and is used in recommending tags. Proposed approaches use L-LDA and similarity-based ranking methods to make tag recommendations. The outcome of this study shows that the proposed approach achieves 0.675 and 0.712 accuracy in terms of recall@10 and recall@5. Chen et al. [23] proposed a semi-supervised tagging method SemiTagRec: a tag recommendation approach for docker repositories. The proposed approach contains four components: Predictor, Extender, Evaluator, and Integrator. The outcome of this study depicts that the proposed approach outperforms STAR [139]and EnTagRec approaches in terms of recall@10 and recall@5 with accuracy 0.781 and 0.688 respectively.

3.2.2 RQ 2.2 What are the various types of smells detected and the approaches used to detect the smell in container images?

To import data and applications during the image building process in containers, temporary files are used. If these temporary files exist in container images, it may lead to an increase in image size and thus effect QoS. These are called temporary file smells. Xu et al. [137] proposed two approaches with static and dynamic analysis to detect temporary file smells. Approach based on dynamic analysis consists of two steps: code injection and log analysis. Approach based on static analysis consists of three steps: building the abstract syntax tree, semantic analysis, and file matching. The result shows that the dynamic approach detects 100% temporary file smell and the static approach is a lightweight and easy-to-use approach. Lu et al. [76] proposed a state-dependent static analysis method to detect temporary files. They also proposed 3 approaches to remove detected temporary file smells:

Just Accepted

instruction merge method, directly copy method, and external storage method. Experimental result shows that the proposed approach effectively and efficiently detect and remove temporary file smells. Wu et al. [135] conducted an empirical study that helps developers to gain some information related to the presence of Docker file smells including its distribution, correlation with project characteristics, coverage, and co-occurrence. They used linear regression to identify the relationship between Docker file smells and project characteristics.

3.2.3 RQ 2.3 What are the image mining approaches used in literature?

Application containers such as docker are gaining popularity these days for software deployment. They can pack irrelevant resources which may lead to an increase in space usage and less security. Rastogi et al. [108] proposed CIMPLIFIER as a solution to detect necessary files in a container which is named as a container image. CIMPLIFIER partitions the container into simple ones and excludes those resources which are not needed to execute the application. CIMPLIFIER follows three steps: resource identification, container partitioning, and gluing. The result of this study depicts that proposed method reduces container image size up to 95% and processes even large containers in less than thirty seconds. As in dynamic analysis, there can be a possibility of inaccuracy in resource set identification because of incomplete code coverage during container execution, and in the case of static analysis approach, there can be an incompatibility in languages. So, Rastogi et al. [109] combine both static and dynamic analysis approaches to build an intermediate representation for container debloating. Civolani et al. [27] proposed FogDocker which reduces the container start-up time on both fog computing server and cloud computing server by finding the necessary files to start the container and deploying these files firstly. The goal of this study is to significantly reduce the time spent needed to deploy the container until the container is ready to perform relevant functionality. Lack of knowledge and experience may lead to the development of Docker files with more quality issues. Hassan et al. [37] proposed RUDSEA which gives information about update locations in Docker files and also recommends changes in docker files based on changes in software environment factors. The outcome of this study depicts that the proposed approach gives correct updates for 44.1% of updates and correct update location for 78.5% of updates.

Discussion

- We discussed about various tagging approaches, various types of code smells and approaches used to detect code smells, and the frameworks introduced to detect the necessary files in a container.
- As there can be a large number of docker repositories that contains docker images, the authors introduced various methods such as D-Tagger, STAR, EnTagRec, and SemiTagRec to assign unique tags to them so that they can be easily retrieved.
- Image detection is the process to identify the necessary files to create the container image so that it consumes less space.

Just Accepted

3.3.1 RQ 3.1 Which type of dataset have been used in container scheduling?

Some authors have used industrial datasets such as online submitted containers requests and some have used open-source datasets such as docker files, nginx, tomcat, debian, ubuntu, alpine, Kubernetes, etc.

Table 6. Dataset Used

Industrial dataset	[78], [89], [90], [91], [92], [94].
Open-source dataset	[5], [11], [32], [38], [39], [43], [49], [52], [61], [62], [63], [66], [69], [82], [86], [88], [106], [115], [117], [121], [122], [123], [124], [125], [129], [133], [153].

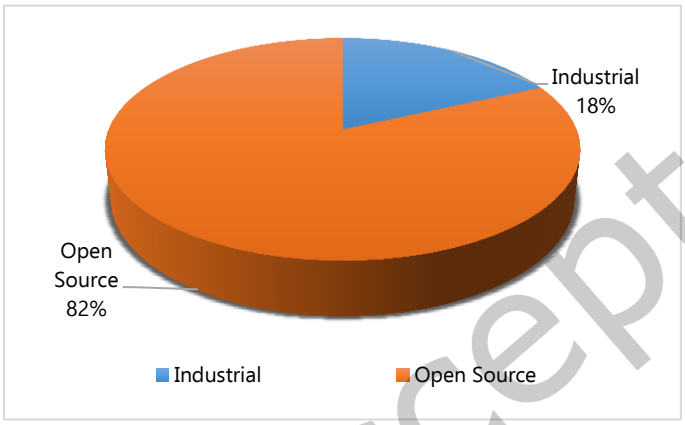


Fig. 4 Dataset Used

Figure 4 describes that in 82% studies, authors have used open-source datasets and in 18% studies, authors have used industrial dataset. Most of the studies in the literature have used open-source dataset. Auver grid trace, google trace data, nginx server images are mostly used open-source dataset in container scheduling techniques.

3.3.2RQ 3.2 What are the various performance objectives used in container scheduling?

Various types of performance objectives such as load balancing, resource utilization, makespan, energy consumption, scalability, cost, etc. have been used in the literature to analyze the performance of container scheduling techniques. An optimization objective is used to access quality of the solution provided by an algorithm. Optimization objectives can be single objective or multi-objective depending on the optimization required for the algorithm [6]. The following performance objectives are frequently used to measure how well the scheduling technique performs:

Just Accepted

Table 7 Performance Objectives Used in Container Scheduling

Performance Objective	Description	Studies	Percentage
Resource Utilization	It describes how efficiently the available resources (memory and network bandwidth) utilized on working nodes. Higher the resource utilization of nodes, the node becomes more energy efficient and cost effective.	[5], [7], [8], [9], [11], [29], [30], [33], [39], [40], [43], [49], [51], [52], [54], [61], [62], [66], [69], [70], [78], [82], [86], [88], [98], [106], [110], [113], [115], [117], [121], [125], [129], [136], [149], [153].	58.06%
Load Balancing	It describes the equal allocation of workload on working nodes so that no working nodes becomes overloaded while other nodes are idle.	[8], [30], [34], [39], [49], [51], [52], [53], [60], [62], [63], [66], [70], [75], [77], [78], [82], [88], [90], [91], [92], [106], [117], [125], [136].	40.32%
Scalability	It deals with the capacity to continuously provide the necessary services despite an increase in load on the system from a single application or from several.	[28], [39], [43], [82], [120].	8.06%
Availability	It indicates the amount of time that a user can access a particular application or service anytime they want to do so.	[7], [8], [30], [33], [38], [43], [89], [123], [133].	14.52%
Energy Consumption	It describes how much energy is consumed when containers are deployed on working nodes. We try to reduce the overall energy consumption rate to lower the computation cost.	[5], [9], [22], [38], [43], [52], [53], [54], [60], [73], [86], [88], [91], [92], [100], [110], [121], [122], [123], [124], [145], [148].	35.48%
Cost	The total cost of an application comprises three aspects: computation, communication, and storage. Computation cost is the time taken to execute an application on available nodes and communication cost is the amount given to network providers for telecommunication services that are required to execute an application.	[29], [32], [73], [78], [99], [115], [129], [145], [153].	14.52%
Makespan	It describes the amount of time needed to execute the application or service from beginning to end. An effective scheduler tries to minimize the overall time.	[5], [9], [11], [30], [34], [40], [54], [60], [63], [70], [75], [78], [90], [91], [99], [111], [113], [115], [133].	30.64%
Network Transmission	It refers to the number of bits transferred per second. This objective helps in identifying communication delay between containers and also helps in managing network congestion.	[33], [66], [82], [110], [111], [145].	9.68%

Just Accepted

3.3.3 RQ 3.3 Which techniques have been used in container scheduling?

Containers are replacing traditional virtual machines as containers share OS, provide fast deployment, and are light weight so they are gaining tremendous popularity nowadays. The problem of efficiently and effectively allocating a node to execute a set of containers is a challenging task. Kubernetes and Docker Swarm orchestration platforms consist of three default container scheduling approaches to deploy the containers on computing nodes. Random strategy schedules the containers on any node randomly, binpack strategy schedules the containers on the node with the most containers, and spread strategy schedules the containers on nodes with the fewest containers. Various scheduling

methodologies have been introduced in the literature. Omogbai Oleghe [102] presented a review of the container scheduling strategy used in edge computing. This study mainly focused on techniques and frameworks that have been used in developing container placement algorithms for edge computing problems. The study demonstrates that optimization modelling frameworks and heuristic-based algorithms are pervasive. Tan et al. [120] compared three heuristic approaches (sub, sum, and random) to show their inability to handle resource requirements well. After this, the author proposed a learning approach named as GPHH to design various rules for container allocation. The result depicts that GPHH outperforms sub, sum, and random approaches in terms of energy consumption. Container scheduling techniques are broadly classified into four categories: Heuristic, Meta- heuristic, Mathematical modelling, and Machine learning.

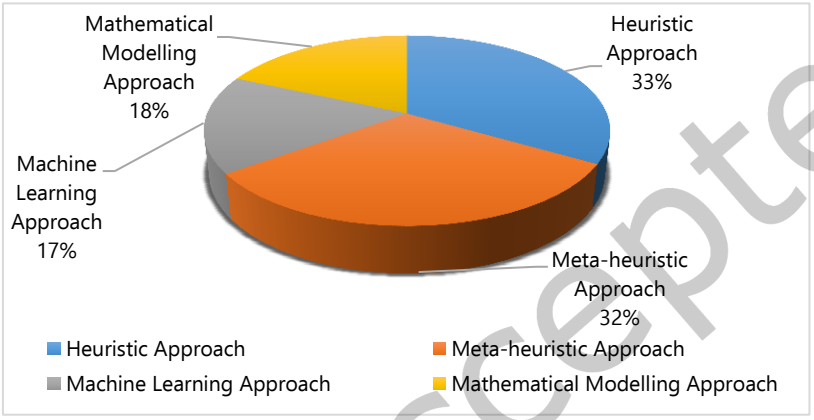


Fig. 5 Scheduling Techniques

Figure 5 shows pie chart of container scheduling techniques. This figure describes that 33% studies are related to heuristic approaches, 32% studies are related to meta-heuristic approaches, 18% studies are related to mathematical modeling approaches, and 17% studies are related to machine learning approaches.

Several heuristic algorithms have been introduced in the literature to solve container scheduling problems. Some of them uses bin packing technique. Most of the proposed approaches use a combination of existing strategies in docker swarm and Kubernetes. Hu et al. [40] introduced ECSched (Enhanced Container Scheduler), an efficient scheduling approach to handle concurrent container requests on heterogeneous clusters. It follows a graph-based approach. This approach first fetches a set of container requests and then applies MCFP (Minimum cost flow problem) algorithm to efficiently place the containers. Experimental result shows that ECSched outperforms docker swarm and kubernetes default scheduling strategy in terms of average container completion time and resource utilization. We have presented a summary of findings related to heuristic approaches in solving container scheduling problems in table 8.

Just Accepted

Table 8 Summary of Studies Related to Heuristic Approaches for Container Scheduling

Studies	Performance Objective	Findings	Techniques Used	Container Used
[89]	Availability	Proposed a new multi-objective scheduling technique which is based on the combination of KUNG and PROMETHEE and used to schedule the containers to a node such that it should consider two objectives: no. of available CPU and memory size.	Multi-objective container scheduling algorithm based on multi-criteria decision analysis.	Docker
[51]	Resource Utilization, Load Balancing	Introduced a multi-dimensional container resource allocation algorithm, CRADLE+ which is designed using a local ratio algorithm. This algorithm allocates nodes or machines to execute the containers to get maximum profit. The result shows that this algorithm achieves 10-50% better results than binpack and spread.	CRADLE+ using local ratio algorithm.	Kubernetes
[90]	Load Balancing, Makespan	Introduced a scheduling strategy by combining spread with binpack using TOPSIS algorithm to select the node that executes container which minimizes the following three criteria: No. of containers in each node, no. of CPUs, and memory size.	Spread+Binpack, TOPSIS (multi-criteria decision algorithm)	Docker
[63]	Load Balance, Mean Squared Error, Makespan	In some cases, swarm not give good performance of load balancing. So, here a new scheduling strategy has been introduced that can meet a various scheduling requirement. Result shows that MCSS strategy performs better in load balancing. The MSE results of MCSS are better than the other strategies. And has an average performance in terms of response time.	Multi algorithm collaboration scheduling strategy (MCSS)	Docker
[29]	Resource Utilization, Cost	To maximize the overall value gained from a workload, a market-value-inspired algorithm for container allocation in Docker Swarm has been proposed. In their scheme, container's execution value varies with time determined by the value curve of the job.	Market-based heuristic approach	Docker
[39]	Load Balancing, Resource Utilization, Scalability	Proposed a container scheduling algorithm with multiple objectives cost functions based on vector bin packing. The objectives include resource usage, load balancing, and network dependency. This algorithm improves overall performance by colocating containers with dependencies on the same host.	Vector bin packing algorithm	Kubernetes

Just Accepted

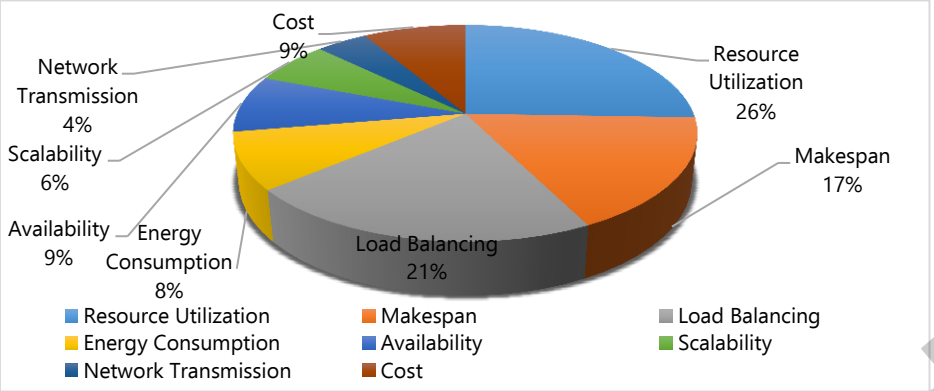


Fig. 6 Performance Objectives Used in Heuristic Approaches

Figure 6 shows pie chart of performance objectives used in heuristic approaches for scheduling the containers on working nodes. This figure describes that among all the performance objectives, resource utilization is the most considered performance objective and network transmission is the least considered performance objective in heuristic approaches. Out of 20 studies, 26% studies used resource utilization, 21% studies used load balancing, 17% studies used makespan, 9% studies used cost, 8% studies used energy consumption, 9% studies used availability, 6% studies used scalability, and only 4% studies used network transmission.

Nowadays, researchers are focusing on metaheuristic algorithms to solve optimization problems in diverse fields. Basically, metaheuristic algorithms are suitable for solving multi-objective optimization problems. Genetic algorithms, particle swarm optimization algorithms, ant colony optimization algorithms, whale optimization algorithms are some

examples of metaheuristic algorithms. Tan et al. [121] proposed new genetic operators, a genetic algorithm to solve two-level container allocation problem, and compared the effectiveness of the proposed GA with best fit descending algorithm and single chromosome GA. The result shows that the proposed algorithm outperforms best fit descending and single chromosome GA in terms of both resource allocation and energy consumption. We have presented a summary of findings related to metaheuristic approaches in solving container scheduling problems in table 9.

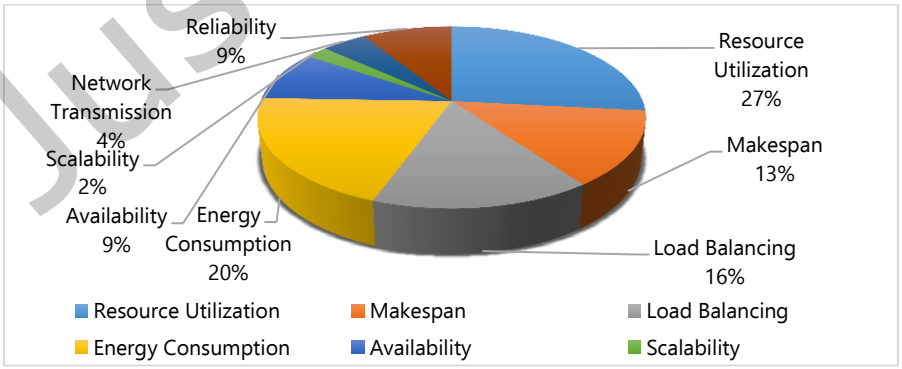


Fig. 7 Performance Objectives Used in Meta Heuristic Approaches

Just Accepted

Table 9. Findings Related to Meta Heuristic Approaches for Container Scheduling

Studies	Performance Objective	Findings	Techniques Used	Container Used
[43]	Availability, Reliability, Scalability, Energy Consumption, Resource Utilization	The result shows that many objective genetic algorithm schedulers improve the result as compared to Ant Colony Optimization algorithm in terms of task distribution, power consumption, unique task assignment, availability, and resource utilization.	Many objective genetic algorithms based on NSGA III algorithm	Docker
[69]	Resource Utilization	The result shows that K-PSO algorithm increases resource utilization by 20% as compared to PSO algorithm, ESS strategy, Kube scheduler default policy, QoS policy.	K-PSO algorithm implemented on Kubernetes platform	Kubernetes, Docker
[61]	Resource Utilization	The result shows that K-PSO algorithm increases resource utilization by 20% as compared to PSO algorithm, ESS strategy, Kube scheduler default policy, QoS policy.	K-PSO algorithm implemented on Kubernetes platform	Docker, Kubernetes
[34]	Load Balancing, Makespan	A container scheduling strategy based on neighbourhood division in micro service (CSBND) is proposed. Compared to the MOPSO algorithm and widely used Spread strategy in micro services, it can increase performance by 20% to 25%.	CSBND integrated with PSO	CloudSim
[62]	Resource Utilization, Load Balancing	Developed particle swarm optimization algorithm to schedule the docker containers to improve resource usage and load balancing. The proposed scheduling approach outperforms spread and random strategies.	PSO (Particle Swarm Optimization) algorithm	Docker
[113]	Makespan, Cost, Resource Utilization, SLA Violations	Used Ant Colony Optimization Algorithm to solve multi objective resource scheduling problem in order to maximize resource utilization with minimizing execution cost, time, SLA violations and space complexity.	Ant Colony Optimization Algorithm	ContainerCloud

Figure 7 shows pie chart of performance objectives used in meta-heuristic approaches for scheduling the containers on working nodes. This figure describes that among all the performance objectives, resource utilization is the most considered performance objective and scalability is the least considered performance objective in meta-heuristic approaches. Out of 19 studies, 27% studies used resource utilization, 10% studies used load balancing, 14% studies used makespan, 22% studies used energy consumption, 10% studies used availability, 5% studies used network transmission, 10% studies used reliability, and only 3% studies used scalability.

Integer Linear Programming (ILP) is a mathematical modelling technique which makes use of a set of linear constraints to optimize linear functions. The primary technique to solve the problems modeled using ILP formulation are branch and bound methods and CPLEX. The ILP formulation of container scheduling problem is NP complete in complexity. Guan et al. [32] developed an application oriented docker container (AODC) framework for resource allocation and formulate resource allocation problem as an optimization problem. The author further proposed a scheduling approach to address AODC optimization problem using linear programming solver (CPLEX). Experimental result shows

Just Accepted

that AODC framework and algorithm outperform worst fit, best fit, and optimal-VM (VM placement by CPLEX) approaches in terms of cost and acceptance ratio. A summary of findings related to mathematical modeling approaches in solving container scheduling problems is presented in table 10.

Table 10. Findings Related to Mathematical Modeling Approaches for Container Scheduling

Studies	Performance Objective	Findings	Techniques Used	Container Used
[52]	Resource Utilization, Energy Consumption, Load Balancing	Introduced Kubernetes based Energy and Interference Driven Scheduler (KEIDS). The proposed scheduler gives an overall improvement of 14.42%, 47%, and 31.83%, respectively against the First come first scheduling technique in terms of energy consumption, Carbon footprint, and load balancing.	KEIDS platform which utilizes Integer linear programming based on multi objective optimization algorithm, Mosek solver for ILP.	Kubernetes
[129]	Resource Utilization, Cost	Proposed a scalable and communication-efficient resource allocation algorithm, named EPTA, to reduce the application deployment cost-constrained by capacity and the service delay bound. To validate the effectiveness, the authors have compared the results of the proposed algorithm with existing approaches in docker swarm and the result depicts that the proposed algorithm is more flexible and less costly.	EPTA (Execution container Placement and Task 18ulnerabili algorithm) which uses linear programming solver (e.g.: CPLEX)	Docker
[53]	Energy Consumption, Load Balancing	Designed multi objective optimization problem (MOOP) which has been expressed as ILP for scheduling containers and proposed an algorithm, EnLoB to solve MOOP to reduce energy consumption and balance the load among nodes.	ILP (Integer Linear Programming), EnLoB (Energy and Load Balancing)	CloudSim
[99]	Cost, Makespan	Proposed EVCD, a general formulation of the container deployment problem for Cloud environments that takes into account the heterogeneity of container requirements and computing resources. The experimental result indicates that EVCD can be used as a benchmark framework against first fit algorithm and round robin algorithm.	EVCD (Elastic provisioning of Virtual machines for Container Deployment)	Docker

Figure 8 shows pie chart of performance objectives used in mathematical modeling approaches for scheduling the containers on working nodes. This figure describes that among all the performance objectives, resource utilization is the most considered performance objective and availability and cost is the least considered performance objective in mathematical modeling approaches. Out of 11 studies, 29% studies used resource utilization, 19% studies used load balancing, 14% studies used makespan, 19% studies used energy consumption, 5% studies used cost, 9% studies used network transmission, and only 5% studies used availability.

Just Accepted

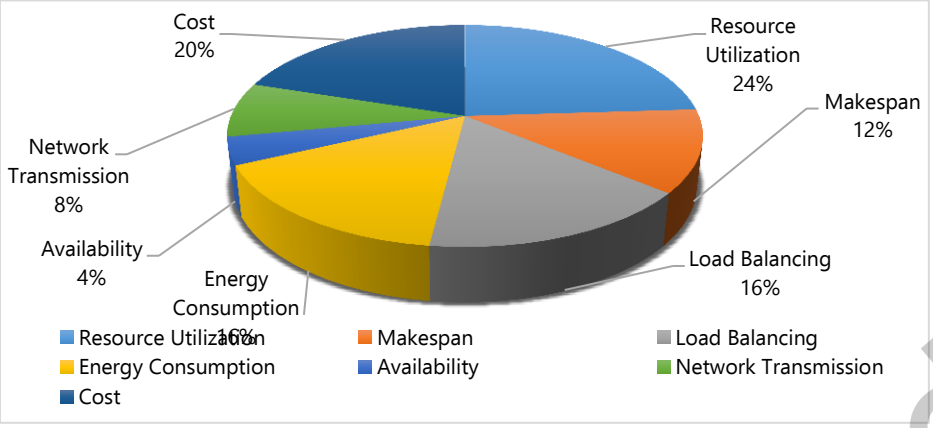


Fig. 8 Performance Objectives Used in Mathematical Modeling Approaches

Machine learning algorithms gather knowledge from data and act accordingly. A subfield of machine learning called Deep learning uses a layered structure of artificial neural networks to learn and make intelligent decisions on its own. Nanda and Hacker [98] formulated a multi-resource bin packing problem and proposed Resource aware container consolidation (RACC) approach, Fit for Packing which uses deep reinforcement learning approach to place containers on physical machine. The result shows that our model achieves an average training accuracy of 82.01% and an average testing accuracy of 82.93% as compared to SJF and random strategy. A summary of findings related to machine learning approaches in solving container scheduling problems is given in table 11.

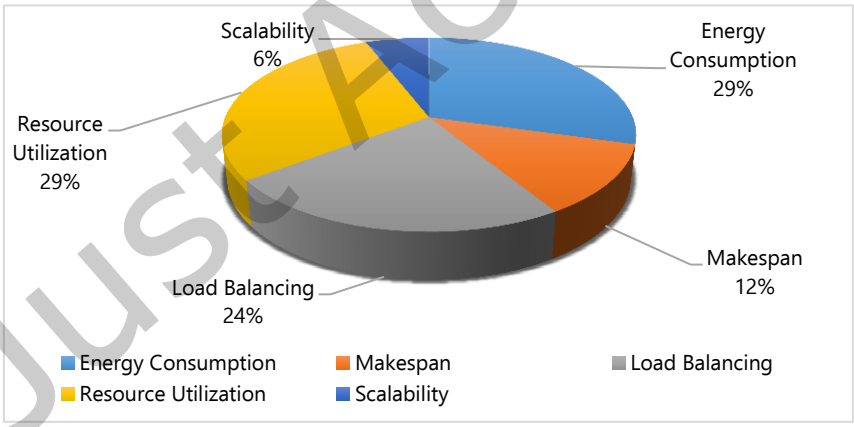


Fig. 9 Performance Objectives Used in Machine Learning Approaches

Just Accepted

Table 11. Findings Related to Machine Learning Approaches for Container Scheduling

Studies	Performance Objective	Findings	Techniques Used	Container Used
[91]	Energy Consumption, Makespan, Load Balancing	This paper addressed the problem of a company that receives containers from their user daily. The idea is to reduce power consumption while satisfying the need of all the users. For this, a new consolidation approach that is based on ML is proposed.	Consolidation approach based on machine learning technique(K-means)	Docker
[92]	Energy Consumption, Load Balancing	A new scheduling technique based on power efficiency is developed by improving the default swarm kit spread strategy. Power consumption gets reduced as compared to default swarm kit spread strategy.	Power consumption based new container scheduling strategy which uses machine learning technique(K-means)	Docker
[77]	Load Balancing	Proposed a new container scheduling technique based on machine learning technique. It maintains load balancing and improves system performance. As compared to the feedback technique, it gives a better result.	Random forest regression technique, Jmeter for load balancing and Prometheus tool to monitor the performance	Kubernetes
[136]	Load Balancing, Resource Utilization	On the basis of container and service mesh technologies, a practical approach is introduced to create a flexible and on-demand system for deep learning applications. And the CPU utilization is 15.52% and the current cluster memory utilization counts for 76% of the overall memory.	Deep Learning, Grafana tool	Docker, Kubernetes

Figure 9 shows pie chart of performance objectives used in machine learning approaches for scheduling the containers on working nodes. This figure describes that among all the performance objectives, resource utilization and energy consumption is the most considered performance objective and scalability is the least considered performance objective in machine learning approaches. Out of 10 studies, 29% studies used resource utilization, 24% studies used load balancing, 29% studies used energy consumption, 12% studies used makespan, and only 6% studies used scalability.

Discussion

- In the analyzed papers, open-source dataset such as google trace, alibaba cluster trace is mostly used.
- Resource utilization, energy consumption, load balancing, makespan, etc. are performance objectives considered in container scheduling approaches. Resource utilization is the most used performance objective.
- Heuristic, meta heuristic, machine learning, and mathematical modeling approaches are used for scheduling containers on working nodes. Heuristic is the most used approach in the analyzed papers.
- Our analysis shows that out of 61 studies related to container scheduling approaches, 33% studies have used heuristic approaches, 32% studies are related to meta heuristic approaches, 18% studies have used mathematical modeling approaches, and 17% studies have used machine learning approaches.

Just Accepted

3.4 RQ 4 What are the various security vulnerabilities, vulnerability detection methods, attacks, and solutions to mitigate the vulnerability issues in containers?

As containers follow kernel sharing property, there is a need to evaluate container security measures. In the literature, authors have identified various security issues and proposed various defense methods and approaches to detect and resolve security vulnerabilities in containers. Lin et al. [67] in their study collected a dataset of 223 exploits from Exploit-db database that may be helpful in attacking containers and categorized them into two parts using two-dimensional taxonomy. First dimension classification is based on the consequences of attacks and second dimension classification is based on the layers at which vulnerable program resides. They also analyzed privilege escalation attacks on containers and proposed a defense method to easily defend the detected privilege escalation attacks. Zerouali et al. [144] presented
an empirical study on outdatedness and vulnerabilities due to third-party packages. They mainly focused on javascript, ruby, and python packages installed in 3000 docker hub community images. These packages may be outdated or missed some releases which lead to potential vulnerabilities in images. The result of this study shows that the maintainers should invest more time and effort in updating third-party packages in order to reduce vulnerabilities in docker hub images. Zerouali et al. [142] analyzed the impact of npm packages in docker images. They analyzed outdatedness and vulnerability of npm packages in docker environment. The result of this study depicts that npm packages are outdated and thus may result in potential security vulnerabilities in docker images. Liu et al. [71] presented an empirical study to understand the security threats in docker environment and identified 3 major sources of security risk in docker containers: sensitive parameters in run commands, malicious docker images, and vulnerabilities in docker images. Authors also suggest solutions to uncover identified threats such as detecting malicious images, automatically resolving vulnerability issues in images, etc.

3.4.1RQ 4.1 What are the various security vulnerabilities in containers and which method have been used to detect the vulnerabilities?

Martin et al. [83] performed vulnerability-oriented analysis to check the security issues of the use of containers in different use cases. They provide a thorough survey of existing studies in the field of docker containers and then made a detailed list of security issues in docker ecosystem. There exists a number of container scanning tools such as AppArmor [154], Anchore [155], Microscanner [156], Clair [157], Cilium [158], etc. which are used to detect the security vulnerabilities in containers. Javed and Toor [44] in their study investigated the quality of various container scanning tools based on coverage and accuracy. The result of their study demonstrates that the existing tool does not detect application package vulnerability and does not have high accuracy. Zerouali et al. [143] presented ConPan, an automated tool to monitor and analyze the outdatedness, vulnerabilities, and bugs. This tool can be used as a CLI or integrated with other processes through its python API.

Just Accepted

Table 12. Findings Related to Approaches to Detect Vulnerabilities

Studies	Approaches Used	Findings
[50]	Proposed deep learning-based algorithm that uses Restricted Boltzmann Machine (RBM)	The result shows that proposed approach protects containers and prevents security hacks of user application and workload in containerized environment.
[57]	Proposed DIVIDS (Docker Image Vulnerability Diagnostic System)	Result shows that DIVIDS provides a reliable docker environment.
[116]	Proposed DIVA (Docker Image Vulnerability Analysis) framework	The result shows that there exist more than 180 vulnerabilities in both official and community images.
[128]	Clair for static analysis, unsupervised algorithms for dynamic analysis	The static method detects 3 vulnerabilities, the dynamic approach detects 22 vulnerabilities and the combined solution of static and dynamic detection approaches detects 24 vulnerabilities out of 28 vulnerabilities.
[151]	Proposed ZeroDVS, a vulnerability detection and container image traceability mechanism	Result shows that ZeroDVS can be deployed in OS in order to trace the images and improve the security of the containers.
[10]	Proposed Docker-sec methodology to protect the containers from outbreaks on container engines and servers. It also secures containers from zero-day vulnerabilities.	The result shows that Docker-sec performs better than existing approaches and provides a security mechanism for docker.
[104]	Introduced machine learning model to statically analyze container images to detect vulnerable python libraries	This approach results in building a user interface to detect vulnerabilities in container images.

3.4.2RQ 4.2 Which methods have been used to resolve the security vulnerability issue?

Sultan et al. [118] presented a literature survey on security issues in containers and the mechanism to secure containers. They have formed four use cases: protect containers from application, protect host from containers, inter container protection, and protect container from a host. The first three use cases utilize software-based solution and the last use case utilize hardware-based solution. Gao et al. [31] identified the container leakage channels that may expose details of host operating systems and containers. In this study, the authors launch advanced power attack named as synergistic power attack. In addition to this, the authors identified the main cause of information leakage and proposed two stage defense approach. The result shows that the proposed method is effective and incurs minor performance overhead. Andreou et al. [74] proposed Docker-sec, an automated mechanism to secure docker containers via the enforcement of access policies with the help of AppArmor. By using Docker-sec, the researchers will learn how Docker-sec can successfully defend containers from zero-day vulnerabilities with low-performance overhead.

3.4.3RQ 4.3 What are the various types of attacks based on attack layers (container, application, docker engine and host) and approaches used to detect security attacks in containerized applications?

Tomar at al. [127] proposed a threat model and presented a detailed classification of attacks based on various layers such as container, host, docker engine and application. Malware, DoS, ARP spoofing, Unauthorized access, etc. take place on the container layer. Poisoned images, outdated software,

Just Accepted

malware, and DoS take place on the allocation layer. Kernel exploit, Cryptojacking, Malicious code injection, etc. take place on the docker engine layer. Kernel exploit, Tampering, Escape attacks, etc. take place on the Host layer. They also provide a case study on the denial-of-service attack in the docker environment. Huang et al. [41] presented an empirical study on the existing docker security mechanism and threats identified in the docker environment so far. They also proposed a threat detection mechanism and used a machine learning technique to predict the unknown attacks in docker images.

Table 13. Findings Related to Approaches to Detect Security Attacks

Studies	Approaches Used	Findings
[47]	Presented CONSERVE framework, multi domain and multi-dimensional framework to select appropriate container security mechanism	The result shows that the presented framework supports an efficient selection of container security monitoring techniques by saving time and effort in matching the techniques with the tasks at hand.
[59]	Proposed SPEAKER (Split Phase Execution of Application Containers)	Experimental result shows that the proposed mechanism incurs less performance overhead and reduces system calls in order to secure containers.
[68]	Proposed CDL (Classified Distributed Learning) framework that uses random forest algorithm for application classification and neural network algorithm for classified learning	The result shows that CDL outperforms simple anomaly detection method as CDL reduces FPR from 12% to 0.24% and increases TPR from 45% to 74%.
[131]	Introduced ContainerGuard, a real time and noise resilient attack detection mechanism	The findings of this study depict that ContainerGuard performs well in detecting meltdown and spectre attacks with an AUC score of 0.9917 and incurs little running performance overhead in container-based big data platforms
[146]	Proposed real time intrusion detection system that is based on OC-SVM (One Class-Support Vector Machine) algorithm	The result of this study demonstrates that OC-SVM detects security attacks with FPR ranging from 0.02 to 0.12 and TPR equal to 1
[132]	Proposed DockerWatch to detect malware in container-based cloud environment. They also proposed CNN and LSTM based detection approach to maintain the tradeoff between high performance and fast real time malware detection.	The result of this study shows that proposed approach efficiently detects malwares with an improvement in runtime performance overhead in container-based cloud environment.

Discussion

- Studies related to various security vulnerabilities in containers, vulnerability detection approaches, approaches to mitigate the vulnerabilities, security attacks, and approaches to detect the security attacks are analyzed.
- There are some open-source tools such as anchore, apparmor, aqua trivy, clair, etc. to detect the vulnerabilities in containers. The authors introduced some frameworks such as DIVA, DIVIDS, DockerSec, ZeroDVS to detect and mitigate the security vulnerabilities.
- Malware, denial of service, malicious code injection, unauthorized access, cryptojacking, etc. are some of the attacks in containers. In the analyzed papers, CONSERVE, SPEAKER, ContainerGuard, CDL, etc. were used for detecting the attacks.
- In the analyzed papers, machine learning and deep learning algorithms were also used to detect the security vulnerabilities and security attacks in containers.

Just Accepted

3.5 RQ 5 Which techniques, benchmarks, and monitoring tools have been used to evaluate the container performance?

In hypervisor-based virtualization, there exist significant performance overhead because each virtual machine runs on its own operating system. Recently, a new prototype of virtualization technology has been introduced which is named as containerization which contains less performance overhead as compared to VM. In this, we will discuss performance evaluation techniques, benchmarks used and monitoring tools used to evaluate container performance. Bachiega et al. [13] presented a survey of studies involving containers and focused on the performance evaluation of containers. They categorized container performance evaluation techniques into three categories: analytical, simulation, and measurement. The result shows that only a few studies used analytical and simulation techniques. Many studies have used benchmarks for performance evaluation which comes under measurement techniques.

3.5.1RQ 5.1 What are the various performance evaluation techniques used to evaluate the performance of the containers?

In the literature, many authors have introduced various frameworks and techniques to evaluate the performance of the containers. Many studies have focused on the performance comparison of hypervisor-based virtualization and containerization approaches. Yoshimura et al. [141] introduced imagejockey framework to continuously evaluate the performance of container images. The experimental results demonstrates that there exist noticeable performance differences due to the deployed environments and features of container images. Liu et al. [72] presented a performance evaluation methodology for industrial applications and then analyzed the results of the proposed methodology by applying it to containerized edge cloud architecture. They also evaluated the processing power of containerizing an edge computing application in terms of implementing neural network and machine learning approaches for industry-based applications. The result shows that the proposed approach can be used by edge-cloud computing users as well as developers to get a client perspective overview of performance of the system.

Manninen et al. [81] analyzed CPU usage, memory usage, and power consumption for various container platforms such as docker, balena, LXC, flatpack for use in Raspberry Pi-based monitoring and controlling device. They used mpstat script to evaluate CPU usage, free script to evaluate memory usage and for power consumption, they used LabJack T7 device. Plauth et al. [105] gave a comparative analysis of performances of unikernels (rumprun, Osv, MirageOS), containers (Docker, LXD), virtualization (KVM, XEN), native hardware, and combinations of these. The result of this study demonstrates that from the application throughput aspect, unikernels perform as equal to containers or better than containers and containers achieve short startup time as compared to unikernels and virtual machines.

Abuabdo and Al-Sharif [2] evaluated the impact of containerization and virtualization techniques on multithreaded applications (written in C and Java). The result shows that in the case of macOS, single-threaded C program gives better results than a multithreaded java program. In the case of windows, single threading in C gives a little bit slower result than single-threaded java program but multithreaded C program outperforms multithreaded java program, and the result for ubuntu and

Just Accepted

docker is same, as docker images are deployed on ubuntu environment. Kennedy Chengeta [24] studied the performance of traditional virtual machines (Xen and KVM) and containers (docker) and also analyzed the performance by combining two technologies (docker and Kubernetes). The result shows that docker containers are faster and take less memory than Xen and KVM. This study concludes that docker with Kubernetes enhances the performance and reduces the delay as well as deployment time.

Table 14 Correlation between Technique and Performance Evaluation Metrics

Studies	Approaches Used	Performance Evaluation Metrics
[72]	Presented a performance evaluation methodology for industrial applications and analyzed the results on containerized edge cloud architecture.	Round trip time, network bandwidth, CPU usage, and memory usage
[141]	Proposed ImageJockey framework for testing the performance of containers. This framework provides benchmark drivers, metric collection, and visualization tools.	CPU usage, throughput, elapsed time, response time, number of page faults, memory usage, etc
[81]	Used mpstat script to evaluate CPU usage, free script to evaluate memory usage and LabJack T7 for power consumption.	CPU usage, memory usage, and power consumption

3.5.2RQ 5.2 Which benchmarks have been used in the performance evaluation of containers?

In this, we consider the findings related to the evaluation of container performance with the help of various benchmark tools. Several benchmark tools have been used in the literature to evaluate container performance such as LINPACK, Y-CRUNCHER, STREAM, NETPERF, BONNIE++, SYSBENCH, etc. Psutil is a benchmark code that is a python library and is used to evaluate CPU utilization, memory utilization, disk utilization, network utilization, sensor utilization. It is mainly useful for system monitoring**,** profiling**,** limiting process resources, and the management of running processe**s**. Ann Mary Joy [48] compared container and virtual machine in terms of application performance and scalability. They used JMeter to evaluate application performance and scalability. The result of this study demonstrates that in terms of scalability and performance, containers outperform virtual machines. Containers can be used for deploying the application to decrease resource overhead because of their better resource utilization and scalability. Benchmark tools are broadly classified into four categories: CPU intensive benchmark, memory intensive benchmark, disk i/o intensive benchmark, and network i/o intensive benchmark.

Just Accepted

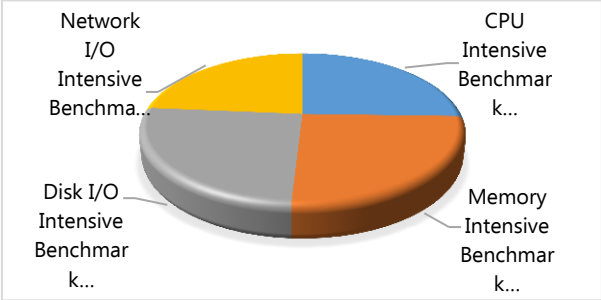


Fig. 10 Benchmark Tools Used

Figure 10 shows pie chart of benchmark tools used for performance evaluation of containers. This figure describes that 26% of the studies used CPU intensive benchmarks, 25% of the studies used memory intensive benchmarks, 25% of the studies used disk i/o intensive benchmark tools, and 24% of the studies used network i/o intensive benchmark tools for performance evaluation. Table 15 describes correlation between benchmark tools and performance evaluation metrics.

Table 15 Correlation between Benchmark Tools and Evaluation Metrics

Studies	Benchmark Tools	Performance Evaluation Metrics
[95], [56], [103], [45], [85], [101], [84], [16], [94], [87], [107], [3], [19], [112]	Y-cruncher, Linpack, Nbench, Noploop, Pbzip2, Geekbench, Sysbench, Spec CPU 2006	CPU usage
[95], [56], [45], [85], [101], [84], [112], [64], [3], [103], [87], [94], [107], [14]	Stream, RAM Speed/SMP	Memory usage
[97], [95], [56], [45], [64], [16], [103], [19], [94], [107], [85], [84], [14], [112]	Bonnie++, Sysbench, Iozine, Fio	Disk i/o usage
[95], [56], [45], [85], [101], [84], [112], [103], [58], [64], [14], [3], [94]	Netperf, Iperf, Nuttcp	Network bandwidth usage

RQ 5.2.1 What are the various CPU intensive benchmarks?

In this category, Y-Cruncher, LINPACK, GEEKBENCH, and SYSBENCH are widely used CPU intensive benchmark tools. Table 16 shows various CPU intensive benchmark tools used in the existing studies.

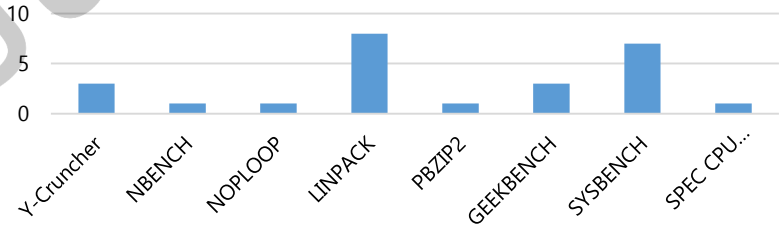


Fig. 11 CPU Intensive Benchmark Tools Used

Just Accepted

Figure 11 shows bar graph of CPU intensive benchmark tools used in the performance evaluation of containers. This figure describes that LINPACK is the most used CPU intensive benchmarks tool with a percentage of 32% and NBENCH, NOPLOOP, PBZIP2, and SPEC CPU 2006 are the least used benchmark tools.

Table 16. CPU Intensive Benchmarks

Benchmark Name	Description	Literature Studies	Percentage
Y-CRUNCHER [159]	Y-cruncher is a multi-threaded and scalable benchmark tool for multi-core systems. It computes Pi and other constants to trillions of digits and is executed as a stress testing tool for the CPU. It produces various outcomes such as total computation time, execution time, and multicore efficiency.	[95],[56],[103]	12%
NBENCH [160]	NBench stands for Native mode Benchmark. It is a single threaded tool used to measure CPU, FPU, and memory speed and it produces three indexes: floating point index, integer index, and memory index.	[95]	4%
NOPLOOP [162]	The *"noploop"* is a very simple CPU benchmarking tool: *"It is a procedure for measuring CPU clock speed using an unrolled No-Operation (NOP) loop. It's intended to be simple, minimizing variation caused by cache misses, stall cycles, and branch misprediction*."	[95]	4%
LINPACK [161]	Linpack is a HPC benchmarking tool used to solve simple linear equations. It performs lower upper decomposition of numerical analysis with partial pivoting. It describes how fast a computer will perform when solving real problems. It generates the outputs in terms of MegaFLOPS (Floating point operations per second).	[95],[56],[45],[85],[101],[84], [16],[94]	32%
PBZIP2 [163]	PBZIP2 is a parallel version of BZIP2 data compression technique which uses pthreads for parallel file compression. It measures the time needed to compress a file. This tool runs faster on multiple CPUs as it firstly divides the file into pieces and then compresses each piece individually.	[56]	4%
GEEKBENCH [164]	Geekbench is a cross platform that analyzes both single core and multi core architectures to compute processor power. Multi core architecture supports the concept of hyperthreading but single core architecture does not support hyperthreading as it uses only one CPU core. It produces various outcomes such as integer performance, memory performance, and floating-point performance.	[56],[16],[87]	12%
SYSBENCH [165]	Sysbench is an open-source benchmark tool used to benchmark Linux. It calculates CPU utilization, memory utilization, disk I/O and database performance. In the case of CPU benchmark, Sysbench evaluates the time required to perform the maximum prime number.	[107],[101],[16],[3], [103],[19],[94]	28%
SPEC CPU 2006[166]	SPEC CPU 2006 is introduced by SPEC (Standard Performance Evaluation Corporation) and stressing memory subsystem, system processor, and compiler. Currently, this benchmark is replaced by SPEC CPU 2017 benchmark suite. It evaluates speed and throughput.	[112]	4%

Just Accepted

RQ 5.2.2 What are the various memory intensive benchmarks?

There exist various benchmark tools to measure memory performance such as STREAM, Graph500, FEniCS, RAM speed/SMP. Graph500 evaluates TEPS (Traversed Edge Per Second) and executes graph data structure to obtain BFS (Breadth First Search) in the data construction and data generation phase. FEniCS solves the Poisson equation and is used in distributed memory parallelism. Table 17 shows various memory intensive benchmark tools used in the existing studies.

Table 17. Memory Intensive Benchmarks

Benchmark Name	Description	Literature Studies	Percentage
STREAM [167]	Stream is an industry standardized memory intensive benchmark that is used to measure sustained memory bandwidth in MB/s. It is also a part of HPC (High Performance Computing) challenge benchmark suite. Stream utilizes four kernels: copy, add, triad and scale and accesses memory throughput.	[95],[56],[45],[85], [101],[84],[112],[64], [3],[103],[87],[94]	86%
RAM speed/SMP [168]	RAM speed/SMP (shared memory multiprocessing or symmetric multiprocessing) is a cache and memory benchmark tool used to measure system memory (both RAM and cache) bandwidth for multiprocessor machines. It consists of two benchmarks (mark and mem): 1. INTmark, FLOATmark, MMXmark, SSEmark and 2. INTmem, FLOATmem, MMXmem, SSEmem.	[107],[14]	14%



Fig. 12 Memory Intensive Benchmark Tools Used

Figure 12 shows bar graph of memory intensive benchmark tools used in the performance evaluation of containers. This figure describes that STREAM is the most used memory intensive benchmarks tool with a percentage of 86% as compared to RAM speed/SMP tool.

RQ 5.2.3 What are the various disk I/O intensive benchmarks?

These types of benchmark tools are used to test disk I/O operations. Tools such as Bonnie++, Iozone, Sysbench, Fio, Dbench, dd test, etc are various disk intensive benchmark tools. Dbench [170] is an emulation of the file system load function of the Netbench benchmark. We created 100 parallel client processes of Dbench to generate load for the file system. Every process begins at the same time and executes the same workload. To test the read and write capacity of the disk system, DD test

Just Accepted

benchmark is used. Table 18 shows various disk I/O intensive benchmark tools used in the existing studies.

Table 18. Disk I/O Intensive Benchmarks

Benchmark Name	Description	Literature Studies	Percentage
Bonnie++ [171]	Bonnie++ is a benchmark suite that is aimed at performing several simple tests of hard drive and file system performance. This tool is used to evaluate the performance of the system based on read/write speed, read/write capability of a system, and also benchmark the number of read/write operations that can be performed per second.	[97],[95],[56],[45],[64],[16],[103]	37%
SYSBENCH [165]	Sysbench is an open-source benchmark tool used to benchmark Linux. It calculates CPU utilization, memory utilization, disk I/O and database performance. To measure file IO performance, we first need to create a test file that is much bigger than your RAM (150GB is a good value).	[56],[16],[103],[19],[94]	26%
IOZONE [169]	IOZONE is a benchmark tool for measuring file I/O performance. To measure file I/O performance, this tool performs various operations such as read, re-read, fread, write, re-write, fwrite, etc.	[107],[85],[84],[14]	21%
FIO [172]	FIO (Flexible Input Output Tester) is a benchmark tool used to measure disk I/O performance by evaluating IOPS (Input/Output Operations Per Second). It has the flexibility to select different IO sizes, sequential or random reads, and writes. It spawns worker threads or processes to perform the specified I/O operations.	[112],[16],[94]	16%

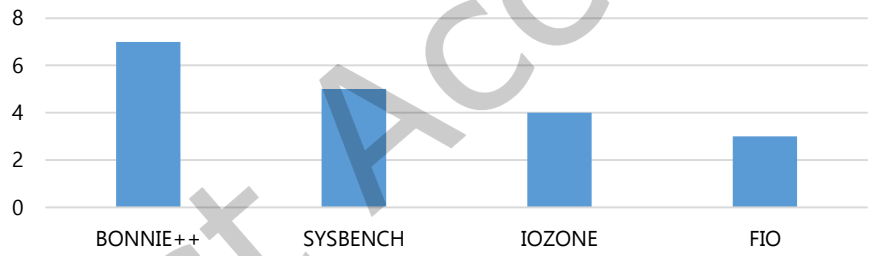


Fig. 13 Disk I/O Intensive Benchmark Tools Used

Figure 13 shows bar graph of disk i/o intensive benchmark tools used in the performance evaluation of containers. This figure describes that BONNIE++ is the most used disk i/o intensive benchmarks tool with a percentage of 37% and FIO is the least used disk i/o intensive benchmark tool.

RQ 5.2.4 What are the various network I/O intensive benchmarks?

There exist various benchmark tools to measure network performance such as Netperf, Iperf, Netpipe, Nuttcp, Iperf3. NetPIPE [173] stands for Network Protocol Independent Performance Evaluator. Its protocol-independent feature shows overhead associated with the protocol layer. It encapsulates the advantages of ttcp and netperf. Iperf [174] is a network intensive benchmark tool used to analyze the network performance in terms of throughput, speed, and bandwidth. This tool creates TCP and UDP

Just Accepted

streams to measure throughput. It follows client server model to measure throughput between two systems. Table 19. shows various network I/O intensive benchmark tools used in the existing studies.

Table 19 Network I/O Intensive Benchmarks

Benchmark Name	Description	Literature Studies	Percentage
NETPERF [175]	This tool measures network bandwidth between two hosts on a network. Netperf is designed on client-server model. There are two executable components in Netperf: netperf (network performance benchmark client) and netserver (network performance benchmark server). It measures request/response performance using TCP, UDP, SCTP.	[95],[56],[45],[85],[101],[84],[112], [103],[58]	53%
IPERF [174]	Iperf is a network intensive benchmark tool used to analyze the network performance in terms of throughput, speed, and bandwidth. This tool creates TCP and UDP streams to measure throughput. It follows client server model to measure throughput between two systems.	[56],[101],[64],[14],[3],[94]	35%
NUTTCP [186]	Nuttcp is a network performance measurement tool that is used by network and system managers. It evaluates throughput in mbps, and also provides other information such as user and system wall clock time, transmitter and receiver CPU utilization, and loss percentage.	[101],[103]	12%



Fig. 14 Network I/O Intensive benchmark tools

Figure 14 shows bar graph of network i/o intensive benchmark tools used in the performance evaluation of containers. This figure describes that NETPERF is the most used network i/o intensive benchmarks tool with a percentage of 53% and NUTTCP is the least used network i/o intensive benchmark tool.

3.5.3 RQ 5.3 Which tools have been used to monitor the container performance?

Container monitoring tools track the operation of containerized applications. Monitoring helps in detecting and resolving the issues at an early stage and ensures optimal system performance all the time. Container monitoring tools assess metrics such as CPU utilization, memory utilization, disk i/o utilization, network bandwidth, CPU limit, and memory limit. [96] presented ConMon tool to monitor the network performance of container-based applications. It is a distributed automated monitoring system that automatically identifies those containers that communicate with each other and analyze the network performance between these containers. The result of this study depicts that ConMon

Just Accepted

system introduces low overhead on resource utilization and negligible overhead on network performance but it can affect throughput and latency of monitored container-based applications. DEEP-mon, a monitoring tool to measure power consumption was proposed by [18]. It is a dynamic and energy-efficient power consumption tool that uses 6 benchmarks EP, MG, CG, pts/apache, pts/nginx, pts/postmark. The experimental result shows that DEEP-mon has a negligible overhead on power consumption and performance. CMonitor, a monitoring and alarming tool for containers proposed by [46]. This tool provides integrated monitoring services, rich visualization to record the data, alarming mechanisms in case of detecting abnormal behavior in containers, and a global topology view. The experimental evaluation depicts that CMonitor can accurately collect CPU utilization, memory utilization, and block I/O metrics. We have presented a summative study of various container performance monitoring tools in table 20.

Table 20 Container Performance Monitoring Tools

Tool name	Logo	Developed by	Visualization tools	Performance metrics	Supported OS	License	Fee	Send Alerts	Target platform
Sematext [176]		Sematext	Kibana, Sematext Logs UI	CPU usage, Disk I/O usage, Memory usage, and Network IO stats	Linux	Open source	Paid	Yes	Docker, rkt
Docker stats [177]		Docker Inc.	No	CPU and memory usage, memory limit, networking information and disk utilization	Linux	Open source	Free	No	Docker
Sensu [178]		Sensu	Grafana	CPU, Memory and network I/O usage	Linux, Windows	Open source, Commercial	Paid	Yes	Docker, LXC,rkt
Sysdig [179]		Sysdig	Sysdig Monitor	CPU, Memory and network I/O usage	Linux, Windows	Open source, Commercial	Paid	Yes	Docker, LXC,rkt
Prometheus [180]		Sound Cloud	Node exporter, Grafana, Prometheus UI	CPU and memory usage and network bandwidth	Linux, Windows, Darwin	Open Source	Free	Yes	Docker
cAdvisor [181]		Google	Grafana, cAdvisor Web UI	CPU and memory usage, network i/o and disk space utilization	Linux	Open source	Free	No	Docker
Datadog [182]		Olivier Pomel, Alexis Lê-Quôc	Dashboard	CPU and disk usage	Window, Linux, Android, iPhone and MacOS	Commercial	Paid	Yes	Docker, Kubernetes

Just Accepted

Scout [183]		Scout	Dashboard	CPU and memory usage	Linux	Open source, Commercial	Paid	Yes	Docker, rkt
Dynatrace [184]		Dynatrace	Smartscape	CPU usage, CPU limit, memory limit and memory usage	Linux, Unix, Windows and z/OS	Open source	Paid	Yes	Docker, Kubernetes, Mesos, Openshift
Appoptics [185]		Solar winds	Dashboard	CPU utilization, memory limits and utilization, network utilization, and error metrics	Linux, MacOS, iPhone	Open source, Commercial	Paid	Yes	Docker, Kubernetes

Discussion

- Our analysis reveals the introduction of various frameworks for assessing container performance along with the utilization of benchmark tools and monitoring tools for performance evaluation of containers.
- The literature offers a list of various CPU intensive (Ycruncher, NBENCH, NOPLOOP, LINPACK, etc.), memory intensive (STREAM, RAM speed/SMP), disk i/o intensive (Bonnie++, Sysbench, IOZONE, etc.), and network i/o intensive (Netperf, Iperf, Nuttcp, etc.) benchmark tools.
- Notably, LINPACK serves as a widely used CPU intensive benchmark tool, STREAM serves as a widely used memory intensive benchmark tool, BONNIE++ serves as a widely used disk i/o intensive benchmark tool, and NETPERF serves as a widely used network i/o intensive benchmark tool.
- Various organizations have also introduced a list of container performance monitoring tools such as cAdvisor, Prometheus, Grafana, docker stats, sematext, dynatrace, etc.

4 CONCLUSION AND FUTURE GUIDELINES

In this paper, the literature on the maintenance of software containers is consolidated by defining a research question framework based on container scheduling techniques, container performance evaluation, container security measures, container performance monitoring tools, image tagging approaches, and image detection approaches. Firstly, we have collected 208 studies from various databases such as IEEE Xplore, Google Scholar, ACM Digital Library, Springer link, Science direct and then we have selected 145 primary studies based on quality assessment criteria and inclusion and exclusion criteria. We have also read 33 hyperlinks to collect the data related to container performance evaluation benchmark tools such as CPU intensive benchmarks tools, memory intensive benchmark tools, disk i/o intensive benchmark tools, and network i/o intensive benchmark tools and container performance monitoring tools. From the summative analysis of selected primary studies on containers, it is found that

Just Accepted

- 42% of studies are related to container scheduling techniques, 24% of studies are related to container performance evaluation, 22% studies are related to container security measures, and 12% studies are related to container image detection process.

- From the perspective of container scheduling algorithms, it is found that 33% studies have used heuristic approaches, 32% studies are related to meta heuristic approaches, 18% studies have used mathematical modeling approaches, and 17% studies have used machine learning approaches.

- In case of heuristic approaches, resource utilization is the most considered performance objective and network transmission is the least considered performance objective. In case of meta heuristic approaches, resource utilization is the most considered performance objective and scalability is the least considered performance objective. In case of mathematical modeling approaches, resource utilization is the most considered performance objective and availability is the least considered performance objective. In case of machine learning approaches, load balancing is the most considered performance objective and makespan is the least considered performance objective.

- Authors have identified various security issues and proposed various defense methods and approaches to detect and resolve security vulnerabilities in containers. Authors have also used various open-source vulnerability scanning tools to detect potential threats and security vulnerabilities in containers.

- Authors have introduced various frameworks to evaluate the performance of containers and have also used benchmark tools for performance evaluation. There exists various CPU intensive, memory intensive, disk i/o intensive, and network i/o intensive benchmark tools in the literature. LINPACK is widely used CPU intensive benchmark tool, STREAM is widely used memory intensive benchmark tool, BONNIE++ is widely used disk i/o intensive benchmark tool, and NETPERF is widely used network i/o intensive benchmark tool.

Since we expect this research area to continue to grow in the future, we hope that our analysis will become useful in developing new approaches for maintaining software containers The following are the future guidelines for researchers to carry out research on maintenance of software containers:

- There are few studies related to container image detection, image tagging approaches, and container security measures. So, there is a need to explore container image detection process and propose new image tagging approaches, new smell detection approaches, and new approaches to detect the necessary images to be included in container.

- There is also a need to explore container security measures and develop new approaches to detect and mitigate security threats as containers are less secure.

- There are few studies that have used machine learning and mathematical modeling approaches in container scheduling. So, in future, we can propose new container scheduling techniques with the help of machine learning techniques and mathematical modeling techniques.

Just Accepted

- As resource utilization is the most considered performance objective in almost all the scheduling approaches so we can analyze the results on the basis of other performance objectives also such as scalability, throughput, network transmission, cost etc.

REFERENCES

[1] Subil Abraham, Arnab K Paul, Redwan Ibne Seraj Khan, and Ali R Butt. 2020. On the Use of Containers in High Performance Computing Environments. In *2020 IEEE 13th International Conference on Cloud Computing (CLOUD)*, 284–293. DOI:https://doi.org/10.1109/CLOUD49709.2020.00048

[2] Amro Abuabdo and Ziad A Al-Sharif. 2019. Virtualization vs. Containerization: Towards a Multithreaded Performance Evaluation Approach. In *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*, 1–6. DOI:https://doi.org/10.1109/AICCSA47632.2019.9035233

[3] Ashijeet Acharya, Jérémy Fanguède, Michele Paolino, and Daniel Raho. 2018. A Performance Benchmarking Analysis of Hypervisors Containers and Unikernels on ARMv8 and x86 CPUs. In *2018 European Conference on Networks and Communications (EuCNC)*, 282–289. DOI:https://doi.org/10.1109/EuCNC.2018.8443248

[4] Jigna N Acharya and Anil C Suthar. 2022. Docker Container Orchestration Management: A Review. In *International Conference on Intelligent Vision and Computing*, 140–153.

[5] Mainak Adhikari and Satish Narayana Srirama. 2019. Multi-objective accelerated particle swarm optimization with a container-based scheduling for Internet-of-Things in cloud environment. *Journal of Network and Computer Applications* 137, (2019), 35–61. DOI:https://doi.org/https://doi.org/10.1016/j.jnca.2019.04.003

[6] Imtiaz Ahmad, Mohammad Gh. AlFailakawi, Asayel AlMutawa, and Latifa Alsalman. 2021. Container scheduling techniques: A Survey and assessment. *Journal of King Saud University - Computer and Information Sciences* (2021). DOI:https://doi.org/https://doi.org/10.1016/j.jksuci.2021.03.002

[7] Yanal Alahmad, Tariq Daradkeh, and Anjali Agarwal. 2018. Availability-Aware Container Scheduler for Application Services in Cloud. In *2018 IEEE 37th International Performance Computing and Communications Conference (IPCCC)*, 1–6. DOI:https://doi.org/10.1109/PCCC.2018.8711295

[8] Yanal Alahmad, Tariq Daradkeh, and Anjali Agarwal. 2019. Optimized Availability-Aware Component Scheduler for Applications in Container-Based Cloud. In *2019 Sixth International Conference on Software Defined Systems (SDS)*, 194–199. DOI:https://doi.org/10.1109/SDS.2019.8768654

[9] Ammar Al-Moalmi, Juan Luo, Ahmad Salah, Kenli Li, and Luxiu Yin. 2021. A whale optimization system for energy-efficient container placement in data centers. *Expert Syst Appl* 164, (2021), 113719. DOI:https://doi.org/https://doi.org/10.1016/j.eswa.2020.113719

[10] Tahir Alyas, Sikandar Ali, Habib Ullah Khan, Ali Samad, Khalid Alissa, and Muhammad Asif Saleem. 2022. Container Performance and Vulnerability Management for Container Security Using Docker Engine. *Security and Communication Networks* 2022, (2022).

[11] A Asensio, X Masip-Bruin, J Garcia, and S Sánchez. 2021. On the optimality of Concurrent Container Clusters Scheduling over heterogeneous smart environments. *Future Generation Computer Systems* 118, (2021), 157–169. DOI:https://doi.org/https://doi.org/10.1016/j.future.2021.01.003

[12] Wissal Attaoui, Essaid Sabir, Halima Elbiaze, and Mohsen Guizani. 2022. VNF and Container Placement: Recent Advances and Future Trends. *arXiv preprint arXiv:2204.00178* (2022).

[13] Naylor G Bachiega, Paulo S L Souza, Sarita M Bruschi, and Simone do R S de Souza. 2018. Container-Based Performance Evaluation: A Survey and Challenges. In *2018 IEEE International Conference on Cloud Engineering (IC2E)*, 398–403. DOI:https://doi.org/10.1109/IC2E.2018.00075

[14] Rabindra K Barik, Rakesh K Lenka, K Rahul Rao, and Devam Ghose. 2016. Performance analysis of virtual machines and containers in cloud computing. In *2016 International Conference on Computing, Communication and Automation (ICCCA)*, 1204–1210. DOI:https://doi.org/10.1109/CCAA.2016.7813925

[15] Ouafa Bentaleb, Adam S Z Belloum, Abderrazak Sebaa, and Aouaouche El-Maouhab. 2022. Containerization technologies: taxonomies, applications and challenges. *J Supercomput* 78, 1 (2022), 1144–1181. DOI:https://doi.org/10.1007/s11227-021-03914-1

[16] Aditya Bhardwaj and C Rama Krishna. 2019. A Container-Based Technique to Improve Virtual Machine Migration in Cloud Computing. *IETE J Res* 0, 0 (2019), 1–16. DOI:https://doi.org/10.1080/03772063.2019.1605848

[17] Aditya Bhardwaj and C Rama Krishna. 2021. Virtualization in Cloud Computing: Moving from Hypervisor to Containerization—A Survey. *Arab J Sci Eng* 46, 9 (2021), 8585–8601. DOI:https://doi.org/10.1007/s13369-021-05553-3

[18] Rolando Brondolin, Tommaso Sardelli, and Marco D Santambrogio. 2018. DEEP-Mon: Dynamic and Energy Efficient Power Monitoring for Container-Based Infrastructures. In *2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, 676–684. DOI:https://doi.org/10.1109/IPDPSW.2018.00110

[19] Emiliano Casalicchio and Vanessa Perciballi. 2017. Measuring Docker Performance: What a Mess!!! In *Proceedings of the 8th ACM/SPEC on International Conference on Performance Engineering Companion* (ICPE '17 Companion), Association for Computing Machinery, New York, NY, USA, 11–16. DOI:https://doi.org/10.1145/3053600.3053605

[20] MinSu Chae, HwaMin Lee, and Kiyeol Lee. 2019. A performance comparison of linux containers and virtual machines using Docker and KVM. *Cluster Comput* 22, 1 (2019), 1765–1775. DOI:https://doi.org/10.1007/s10586-017-1511-2

Under Accepted

[21] Ngoc-Tu Chau, Jaehyeon Yoon, Thien-Phuc Doan, and Souhwan Jung. 2021. AppPACK: A Packaging Model for Single-Purpose Lightweight Virtualization Environment. *IEEE Access* 9, (2021), 30071–30079.

[22] Feifei Chen, Xiaofeng Zhou, and Chao Shi. 2019. The Container Scheduling Method Based on the Min-Min in Edge Computing. In *Proceedings of the 2019 4th International Conference on Big Data and Computing* (ICBDC 2019), Association for Computing Machinery, New York, NY, USA, 83–90. DOI:https://doi.org/10.1145/3335484.3335506

[23] Wei Chen, Jia-Hong Zhou, Jia-Xin Zhu, Guo-Quan Wu, and Jun Wei. 2019. Semi-supervised learning based tag recommendation for docker repositories. *J Comput Sci Technol* 34, 5 (2019), 957–971.

[24] Kennedy Chengeta. 2021. Comparing the Performance between Virtual Machines and Containers Using Deep Learning Credit Models. In *Proceedings of the International Conference on Artificial Intelligence and Its Applications* (icARTi '21), Association for Computing Machinery, New York, NY, USA. DOI:https://doi.org/10.1145/3487923.3487934

[25] Ron C Chiang. 2020. Contention-aware container placement strategy for docker swarm with machine learning based clustering algorithms. *Cluster Comput* (2020), 1–11.

[26] Jürgen Cito, Gerald Schermann, John Erik Wittern, Philipp Leitner, Sali Zumberi, and Harald C Gall. 2017. An empirical analysis of the docker container ecosystem on github. In *2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR)*, 323–333.

[27] Lorenzo Civolani, Guillaume Pierre, and Paolo Bellavista. 2019. FogDocker: Start container now, fetch image later. In *Proceedings of the 12th IEEE/ACM International Conference on Utility and Cloud Computing*, 51–59.

[28] Qiqing Deng, Xinrui Tan, Jing Yang, Chao Zheng, Liming Wang, and Zhen Xu. 2022. A Secure Container Placement Strategy Using Deep Reinforcement Learning in Cloud. In *2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 1299–1304.

[29] Piotr Dziurzanski and Leandro Soares Indrusiak. 2018. Value-Based Allocation of Docker Containers. In *2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, 358–362. DOI:https://doi.org/10.1109/PDP2018.2018.00064

[30] Guisheng Fan, Liang Chen, Huiqun Yu, and Wei Qi. 2020. Multi-objective optimization of container-based microservice scheduling in edge computing. *Computer Science and Information Systems 18*, 23–42. DOI:https://doi.org/10.2298/CSIS200229041F

[31] Xing Gao, Zhongshu Gu, Mehmet Kayaalp, Dimitrios Pendarakis, and Haining Wang. 2017. ContainerLeaks: Emerging Security Threats of Information Leakages in Container Clouds. In *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 237–248. DOI:https://doi.org/10.1109/DSN.2017.49

[32] Xinjie Guan, Xili Wan, Baek-Young Choi, Sejun Song, and Jiafeng Zhu. 2017. Application Oriented Dynamic Resource Allocation for Data Centers Using Docker Containers. *IEEE Communications Letters* 21, 3 (2017), 504–507. DOI:https://doi.org/10.1109/LCOMM.2016.2644658

[33] Carlos Guerrero, Isaac Lera, and Carlos Juiz. 2018. Genetic algorithm for multi-objective optimization of container allocation in cloud architecture. *J Grid Comput* 16, 1 (2018), 113–135.

[34] Yanghu Guo and Wenbin Yao. 2018. A container scheduling strategy based on neighborhood division in micro service. In *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, 1–6. DOI:https://doi.org/10.1109/NOMS.2018.8406285

[35] Mubin Ul Haque and M Ali Babar. 2022. Well begun is half done: An empirical study of exploitability \& impact of base-image vulnerabilities. In *2022 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*, 1066–1077.

[36] Sanjay Hardikar, Pradeep Ahirwar, and Sameer Rajan. 2021. Containerization: Cloud Computing based Inspiration Technology for Adoption through Docker and Kubernetes. In *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 1996–2003. DOI:https://doi.org/10.1109/ICESC51422.2021.9532917

[37] Foyzul Hassan, Rodney Rodriguez, and Xiaoyin Wang. 2018. Rudsea: recommending updates of dockerfiles via software environment analysis. In *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering*, 796–801.

[38] Aurelien Havet, Valerio Schiavoni, Pascal Felber, Maxime Colmant, Romain Rouvoy, and Christof Fetzer. 2017. Genpack: A generational scheduler for cloud data centers. In *2017 IEEE International Conference on Cloud Engineering (IC2E)*, 95–104.

[39] Yang Hu, Cees de Laat, and Zhiming Zhao. 2019. Multi-objective container deployment on heterogeneous clusters. In *2019 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, 592–599.

[40] Yang Hu, Huan Zhou, Cees de Laat, and Zhiming Zhao. 2020. Concurrent container scheduling on heterogeneous clusters with multi-resource constraints. *Future Generation Computer Systems* 102, (2020), 562–573. DOI:https://doi.org/https://doi.org/10.1016/j.future.2019.08.025

[41] Delu Huang, Handong Cui, Shihao Wen, and Cheng Huang. 2019. Security Analysis and Threats Detection Techniques on Docker Container. In *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*, 1214–1220. DOI:https://doi.org/10.1109/ICCC47050.2019.9064441

[42] Md Hasan Ibrahim, Mohammed Sayagh, and Ahmed E Hassan. 2020. Too many images on DockerHub! How different are images for the same system? *Empir Softw Eng* 25, 5 (2020), 4250–4281.

[43] Mahmoud Imdoukh, Imtiaz Ahmad, and Mohammad Alfailakawi. 2020. Optimizing scheduling decisions of container management tool using many-objective genetic algorithm. *Concurr Comput* 32, 5 (2020), e5536. DOI:https://doi.org/https://doi.org/10.1002/cpe.5536

[44] Omar Javed and Salman Toor. 2021. Understanding the Quality of Container Security Vulnerability Detection Tools. DOI:https://doi.org/10.48550/ARXIV.2101.03844

Just Accepted

[45] Devki Nandan Jha, Saurabh Garg, Prem Prakash Jayaraman, Rajkumar Buyya, Zheng Li, and Rajiv Ranjan. 2018. A Holistic Evaluation of Docker Containers for Interfering Microservices. In *2018 IEEE International Conference on Services Computing (SCC)*, 33–40. DOI:https://doi.org/10.1109/SCC.2018.00012

[46] Shujian Ji, Kejiang Ye, and Cheng-Zhong Xu. 2019. CMonitor: A Monitoring and Alarming Platform for Container-Based Clouds. In *Cloud Computing -- CLOUD 2019*, Springer International Publishing, Cham, 324–339.

[47] Rodi Jolak, Thomas Rosenstatter, Mazen Mohamad, Kim Strandberg, Behrooz Sangchoolie, Nasser Nowdehi, and Riccardo Scandariato. 2022. CONSERVE: A framework for the selection of techniques for monitoring containers security. *Journal of Systems and Software* 186, (2022), 111158. DOI:https://doi.org/https://doi.org/10.1016/j.jss.2021.111158

[48] Ann Mary Joy. 2015. Performance comparison between Linux containers and virtual machines. In *2015 International Conference on Advances in Computer Engineering and Applications*, 342–346. DOI:https://doi.org/10.1109/ICACEA.2015.7164727

[49] Chanwit Kaewkasi and Kornrathak Chuenmuneewong. 2017. Improvement of container scheduling for Docker using Ant Colony Optimization. In *2017 9th International Conference on Knowledge and Smart Technology (KST)*, 254–259. DOI:https://doi.org/10.1109/KST.2017.7886112

[50] Supriya Kamthania. 2019. A Novel Deep Learning RBM Based Algorithm for Securing Containers. In *2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, 1–7. DOI:https://doi.org/10.1109/WIECON-ECE48653.2019.9019985

[51] Oren Katz, Dror Rawitz, and Danny Raz. 2021. Containers Resource Allocation in Dynamic Cloud Environments. In *2021 IFIP Networking Conference (IFIP Networking)*, 1–9. DOI:https://doi.org/10.23919/IFIPNetworking52078.2021.9472812

[52] Kuljeet Kaur, Sahil Garg, Georges Kaddoum, Syed Hassan Ahmed, and Mohammed Atiquzzaman. 2020. KEIDS: Kubernetes-Based Energy and Interference Driven Scheduler for Industrial IoT in Edge-Cloud Ecosystem. *IEEE Internet Things J* 7, 5 (2020), 4228–4237. DOI:https://doi.org/10.1109/JIOT.2019.2939534

[53] Kuljeet Kaur, Sahil Garg, Georges Kaddoum, Francois Gagnon, and Dushantha Nalin K Jayakody. 2019. EnLoB: Energy and Load Balancing-Driven Container Placement Strategy for Data Centers. In *2019 IEEE Globecom Workshops (GC Wkshps)*, 1–6. DOI:https://doi.org/10.1109/GCWkshps45667.2019.9024592

[54] Saad Ahmad Khan, Muhammad Abdullah, Waheed Iqbal, Muhammad Arif Butt, Faisal Bukhari, and Saeed-Ul Hassan. 2022. Automatic Migration-Enabled Dynamic Resource Management for Containerized Workload. *IEEE Syst J* (2022).

[55] Barbara Ann Kitchenham and Stuart Charters. 2007. *Guidelines for performing Systematic Literature Reviews in Software Engineering*. Retrieved from https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf

[56] Zhanibek Kozhirbayev and Richard O Sinnott. 2017. A performance comparison of container-based technologies for the Cloud. *Future Generation Computer Systems* 68, (2017), 175–182. DOI:https://doi.org/https://doi.org/10.1016/j.future.2016.08.025

[57] Soonhong Kwon and Jong-Hyouk Lee. 2020. DIVDS: Docker Image Vulnerability Diagnostic System. *IEEE Access* 8, (2020), 42666–42673. DOI:https://doi.org/10.1109/ACCESS.2020.2976874

[58] Kyungwoon Lee, Hyunjin Kim, Beomsu Kim, and Chuck Yoo. 2017. Analysis on network performance of container virtualization on IoT devices. In *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, 35–37. DOI:https://doi.org/10.1109/ICTC.2017.8190937

[59] Lingguang Lei, Jianhua Sun, Kun Sun, Chris Shenefiel, Rui Ma, Yuewu Wang, and Qi Li. 2017. SPEAKER: Split-phase execution of application containers. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, 230–251.

[60] Feng Li, Wen Jun Tan, and Wentong Cai. 2022. A wholistic optimization of containerized workflow scheduling and deployment in the cloud--edge environment. *Simul Model Pract Theory* 118, (2022), 102521.

[61] Jiawei Li, Bo Liu, Weiwei Lin, Pengfei Li, and Qian Gao. 2019. An Improved Container Scheduling Algorithm Based on PSO for Big Data Applications. In *Cyberspace Safety and Security*, Springer International Publishing, Cham, 516–530.

[62] Lianwan Li, Jianxin Chen, and Wuyang Yan. 2018. A Particle Swarm Optimization-Based Container Scheduling Algorithm of Docker Platform. In *Proceedings of the 4th International Conference on Communication and Information Processing* (ICCIP '18), Association for Computing Machinery, New York, NY, USA, 12–17. DOI:https://doi.org/10.1145/3290420.3290432

[63] Qilong Li and Yu Fang. 2017. Multi-Algorithm Collaboration Scheduling Strategy for Docker Container. In *2017 International Conference on Computer Systems, Electronics and Control (ICCSEC)*, 1367–1371.

[64] Zheng Li, Maria Kihl, Qinghua Lu, and Jens A Andersson. 2017. Performance Overhead Comparison between Hypervisor and Container Based Virtualization. In *2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*, 955–962. DOI:https://doi.org/10.1109/AINA.2017.79

[65] Changyuan Lin, Sarah Nadi, and Hamzeh Khazaei. 2020. A large-scale data set and an empirical study of docker images hosted on docker hub. In *2020 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, 371–381.

[66] Miao Lin, Jianqing Xi, Weihua Bai, and Jiayin Wu. 2019. Ant Colony Algorithm for Multi-Objective Optimization of Container-Based Microservice Scheduling in Cloud. *IEEE Access* 7, (2019), 83088–83100. DOI:https://doi.org/10.1109/ACCESS.2019.2924414

[67] Xin Lin, Lingguang Lei, Yuewu Wang, Jiwu Jing, Kun Sun, and Quan Zhou. 2018. A Measurement Study on Linux Container Security: Attacks and Countermeasures. In *Proceedings of the 34th Annual Computer Security Applications Conference* (ACSAC '18), Association for Computing Machinery, New York, NY, USA, 418–429. DOI:https://doi.org/10.1145/3274694.3274720

[68] Yuhang Lin, Olufogorehan Tunde-Onadele, and Xiaohui Gu. 2020. CDL: Classified Distributed Learning for Detecting Security Attacks in Containerized Applications. In *Annual Computer Security Applications Conference* (ACSAC '20), Association for Computing Machinery, New York, NY, USA, 179–188. DOI:https://doi.org/10.1145/3427228.3427236

Just Accepted

[69] Bo Liu, Jiawei Li, Weiwei Lin, Weihua Bai, Pengfei Li, and Qian Gao. 2021. K-PSO: An improved PSO-based container scheduling algorithm for big data applications. *International Journal of Network Management* 31, 2 (2021), e2092. DOI:https://doi.org/https://doi.org/10.1002/nem.2092

[70] Bo Liu, Pengfei Li, Weiwei Lin, Na Shu, Yin Li, and Victor Chang. 2018. A new container scheduling algorithm based on multi-objective optimization. *Soft comput* 22, 23 (2018), 7741–7752.

[71] Peiyu Liu, Shouling Ji, Lirong Fu, Kangjie Lu, Xuhong Zhang, Wei-Han Lee, Tao Lu, Wenzhi Chen, and Raheem Beyah. 2020. Understanding the Security Risks of Docker Hub. In *Computer Security -- ESORICS 2020*, Springer International Publishing, Cham, 257–276.

[72] Yu Liu, Dapeng Lan, Zhibo Pang, Magnus Karlsson, and Shaofang Gong. 2021. Performance Evaluation of Containerization in Edge-Cloud Computing Stacks for Industrial Applications: A Client Perspective. *IEEE Open Journal of the Industrial Electronics Society* 2, (2021), 153–168. DOI:https://doi.org/10.1109/OJIES.2021.3055901

[73] Saiqin Long, Wen Wen, Zhetao Li, Kenli Li, Rong Yu, and Jiang Zhu. 2021. A Global Cost-aware Container Scheduling Strategy in Cloud Data Centers. *IEEE Transactions on Parallel & Distributed Systems* 01 (2021), 1.

[74] Fotis Loukidis-Andreou, Ioannis Giannakopoulos, Katerina Doka, and Nectarios Koziris. 2018. Docker-Sec: A Fully Automated Container Security Enhancement Mechanism. In *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, 1561–1564. DOI:https://doi.org/10.1109/ICDCS.2018.00169

[75] Weizhi Lu, Boyu Li, and Bin Wu. 2019. Overhead Aware Task Scheduling for Cloud Container Services. In *2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 380–385. DOI:https://doi.org/10.1109/CSCWD.2019.8791871

[76] Zhigang Lu, Jiwei Xu, Yuewen Wu, Tao Wang, and Tao Huang. 2019. An empirical case study on the temporary file smell in dockerfiles. *IEEE Access* 7, (2019), 63650–63659.

[77] Jingze Lv, Mingchang Wei, and Yang Yu. 2019. A Container Scheduling Strategy Based on Machine Learning in Microservice Architecture. In *2019 IEEE International Conference on Services Computing (SCC)*, 65–71. DOI:https://doi.org/10.1109/SCC.2019.00023

[78] Liang Lv, Yuchao Zhang, Yusen Li, Ke Xu, Dan Wang, Wendong Wang, Minghui Li, Xuan Cao, and Qingqing Liang. 2019. Communication-aware container placement and reassignment in large-scale internet data centers. *IEEE Journal on Selected Areas in Communications* 37, 3 (2019), 540–555.

[79] R Madhumathi. 2018. The Relevance of Container Monitoring Towards Container Intelligence. In *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–5. DOI:https://doi.org/10.1109/ICCCNT.2018.8493766

[80] Ruchika Malhotra. 2016. *Empirical research in software engineering: concepts, analysis, and applications*. CRC press.

[81] Harri Manninen, Vesa Jääskeläinen, and Jan Olaf Blech. 2020. Performance Evaluation of Containerization Platforms for Control and Monitoring Devices. In *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 1061–1064. DOI:https://doi.org/10.1109/ETFA46521.2020.9211901

[82] Ying Mao, Jenna Oak, Anthony Pompili, Daniel Beer, Tao Han, and Peizhao Hu. 2017. Draps: Dynamic and resource-aware placement scheme for docker containers in a heterogeneous cluster. In *2017 IEEE 36th International Performance Computing and Communications Conference (IPCCC)*, 1–8.

[83] A Martin, S Raponi, T Combe, and R di Pietro. 2018. Docker ecosystem – Vulnerability Analysis. *Comput Commun* 122, (2018), 30–43. DOI:https://doi.org/https://doi.org/10.1016/j.comcom.2018.03.011

[84] Ilias Mavridis and Helen Karatza. 2017. Performance and Overhead Study of Containers Running on Top of Virtual Machines. In *2017 IEEE 19th Conference on Business Informatics (CBI)*, 32–38. DOI:https://doi.org/10.1109/CBI.2017.69

[85] Ilias Mavridis and Helen Karatza. 2019. Combining containers and virtual machines to enhance isolation and extend functionality on cloud computing. *Future Generation Computer Systems* 94, (2019), 674–696. DOI:https://doi.org/https://doi.org/10.1016/j.future.2018.12.035

[86] Hemant Kumar Mehta, Paul Harvey, Omer Rana, Rajkumar Buyya, and Blesson Varghese. 2020. WattsApp: Power-Aware Container Scheduling. In *2020 IEEE/ACM 13th International Conference on Utility and Cloud Computing (UCC)*, 79–90. DOI:https://doi.org/10.1109/UCC48980.2020.00027

[87] Tianhui Meng, Zhihao Shang, and Katinka Wolter. 2017. An empirical performance and security evaluation of android container solutions. In *2017 International Conference on Cyber Security And Protection Of Digital Services (Cyber Security)*, 1–8. DOI:https://doi.org/10.1109/CyberSecPODS.2017.8074850

[88] Tarek Menouer. 2021. KCSS: Kubernetes container scheduling strategy. *J Supercomput* 77, 5 (2021), 4267–4293.

[89] Tarek Menouer, Christophe Cérin, and Étienne Leclercq. 2018. New multi-objectives scheduling strategies in docker swarmkit. In *International Conference on Algorithms and Architectures for Parallel Processing*, 103–117.

[90] Tarek Menouer and Patrice Darmon. 2019. New scheduling strategy based on multi-criteria decision algorithm. In *2019 27th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, 101–107.

[91] Tarek Menouer and Patrice Darmon. 2019. Containers scheduling consolidation approach for cloud computing. In *International Symposium on Pervasive Systems, Algorithms and Networks*, 178–192.

[92] Tarek Menouer, Otman Manad, Christophe Cérin, and Patrice Darmon. 2019. Power efficiency containers scheduling approach based on machine learning technique for cloud computing environment. In *International Symposium on Pervasive Systems, Algorithms and Networks*, 193–206.

[93] Alan Mills, Jonathan White, and Phil Legg. 2022. OGMA: Visualisation for Software Container Security Analysis and Automated Remediation. In *2022 IEEE International Conference on Cyber Security and Resilience (CSR)*, 76–81.

Just Accepted

[94] Roberto Morabito. 2017. Virtualization on Internet of Things Edge Devices With Container Technologies: A Performance Evaluation. *IEEE Access* 5, (2017), 8835–8850. DOI:https://doi.org/10.1109/ACCESS.2017.2704444

[95] Roberto Morabito, Jimmy Kjällman, and Miika Komu. 2015. Hypervisors vs. Lightweight Virtualization: A Performance Comparison. In *2015 IEEE International Conference on Cloud Engineering*, 386–393. DOI:https://doi.org/10.1109/IC2E.2015.74

[96] Farnaz Moradi, Christofer Flinta, Andreas Johnsson, and Catalin Meirosu. 2017. ConMon: An automated container based network performance monitoring system. In *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, 54–62. DOI:https://doi.org/10.23919/INM.2017.7987264

[97] Preeth E N, Fr. Jaison Paul Mulerickal, Biju Paul, and Yedhu Sastri. 2015. Evaluation of Docker containers based on hardware utilization. In *2015 International Conference on Control Communication Computing India (ICCC)*, 697–700. DOI:https://doi.org/10.1109/ICCC.2015.7432984

[98] Saurav Nanda and Thomas J Hacker. 2018. RACC: Resource-Aware Container Consolidation Using a Deep Learning Approach. In *Proceedings of the First Workshop on Machine Learning for Computing Systems* (MLCS'18), Association for Computing Machinery, New York, NY, USA. DOI:https://doi.org/10.1145/3217871.3217876

[99] Matteo Nardelli, Christoph Hochreiner, and Stefan Schulte. 2017. Elastic Provisioning of Virtual Machines for Container Deployment. In *Proceedings of the 8th ACM/SPEC on International Conference on Performance Engineering Companion* (ICPE '17 Companion), Association for Computing Machinery, New York, NY, USA, 5–10. DOI:https://doi.org/10.1145/3053600.3053602

[100] Shubha Brata Nath, Sourav Kanti Addya, Sandip Chakraborty, and Soumya K Ghosh. 2020. Green Containerized Service Consolidation in Cloud. In *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 1–6. DOI:https://doi.org/10.1109/ICC40277.2020.9149173

[101] Vivian Noronha, Ekkehard Lang, Maximilian Riegel, and Thomas Bauschert. 2018. Performance Evaluation of Container Based Virtualization on Embedded Microprocessors. In *2018 30th International Teletraffic Congress (ITC 30)*, 79–84. DOI:https://doi.org/10.1109/ITC30.2018.00019

[102] Omogbai Oleghe. 2021. Container Placement and Migration in Edge Computing: Concept and Scheduling Models. *IEEE Access* 9, (2021), 68028–68043. DOI:https://doi.org/10.1109/ACCESS.2021.3077550

[103] Arnaldo Pereira Ferreira and Richard Sinnott. 2019. A Performance Evaluation of Containers Running on Managed Kubernetes Services. In *2019 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, 199–208. DOI:https://doi.org/10.1109/CloudCom.2019.00038

[104] Jayama Pinnamaneni, S Nagasundari, and Prasad Honnavalli. 2022. Identifying Vulnerabilities in Docker Image Code using ML Techniques. In *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)*, 1–5.

[105] Max Plauth, Lena Feinbube, and Andreas Polze. 2017. A Performance Survey of Lightweight Virtualization Techniques. In *Service-Oriented and Cloud Computing*, Springer International Publishing, Cham, 34–48.

[106] U Pongsakorn, Yasuhiro Watashiba, Kohei Ichikawa, Susumu Date, Hajimu Iida, and others. 2017. Container rebalancing: Towards proactive linux containers placement optimization in a data center. In *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, 788–795.

[107] Amit M Potdar, Narayan D G, Shivaraj Kengond, and Mohammed Moin Mulla. 2020. Performance Evaluation of Docker Container and Virtual Machine. *Procedia Comput Sci* 171, (2020), 1419–1428. DOI:https://doi.org/https://doi.org/10.1016/j.procs.2020.04.152

[108] Vaibhav Rastogi, Drew Davidson, Lorenzo de Carli, Somesh Jha, and Patrick McDaniel. 2017. Cimplifier: automatically debloating containers. In *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering*, 476–486.

[109] Vaibhav Rastogi, Chaitra Niddodi, Sibin Mohan, and Somesh Jha. 2017. New directions for container debloating. In *Proceedings of the 2017 Workshop on Forming an Ecosystem Around Software Transformation*, 51–56.

[110] Leonardo R Rodrigues, Marcelo Pasin, Omir C Alves, Charles C Miers, Mauricio A Pillon, Pascal Felber, and Guilherme P Koslovski. 2019. Network-Aware Container Scheduling in Multi-Tenant Data Center. In *2019 IEEE Global Communications Conference (GLOBECOM)*, 1–6. DOI:https://doi.org/10.1109/GLOBECOM38437.2019.9013128

[111] Leonardo Rosa Rodrigues, Guilherme Piêgas Koslovski, Marcelo Pasin, Maurício Aronne Pillon, Omir Correia Alves, and Charles Christian Miers. 2021. Time-constrained and network-aware containers scheduling in GPU era. *Future Generation Computer Systems* 117, (2021), 72–86. DOI:https://doi.org/https://doi.org/10.1016/j.future.2020.11.014

[112] Bowen Ruan, Hang Huang, Song Wu, and Hai Jin. 2016. A Performance Study of Containers in Cloud Environment. In *Advances in Services Computing*, Springer International Publishing, Cham, 343–356.

[113] Mufeed Ahmed Naji Saif, Abhinava Karantha, S K Niranjan, and Belal Abdullah Hezam Murshed. 2022. Multi Objective Resource Scheduling for Cloud Environment using Ant Colony Optimization Algorithm. *J Algebr Stat* 13, 3 (2022), 2798–2809.

[114] A Sergeev, E Rezedinova, and A Khakhina. 2022. Stress testing of Docker containers running on a Windows operating system. In *Journal of Physics: Conference Series*, 12010.

[115] Zhenwu Shi, Chenming Jiang, Landu Jiang, and Xue Liu. 2021. HPKS: High Performance Kubernetes Scheduling for Dynamic Blockchain Workloads in Cloud Computing. In *2021 IEEE 14th International Conference on Cloud Computing (CLOUD)*, 456–466.

[116] Rui Shu, Xiaohui Gu, and William Enck. 2017. A Study of Security Vulnerabilities on Docker Hub. In *Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy* (CODASPY '17), Association for Computing Machinery, New York, NY, USA, 269–280. DOI:https://doi.org/10.1145/3029806.3029832

Just Accepted

[117] Shengbo Song, Lelai Deng, Jun Gong, and Hanmei Luo. 2018. Gaia scheduler: A kubernetes-based scheduler framework. In *2018 IEEE Intl Conf on Parallel \& Distributed Processing with Applications, Ubiquitous Computing \& Communications, Big Data \& Cloud Computing, Social Computing \& Networking, Sustainable Computing \& Communications (ISPA/IUCC/BDCloud/SocialCom/SustainC*, 252–259.

[118] Sari Sultan, Imtiaz Ahmad, and Tassos Dimitriou. 2019. Container Security: Issues, Challenges, and the Road Ahead. *IEEE Access* 7, (2019), 52976–52996. DOI:https://doi.org/10.1109/ACCESS.2019.2911732

[119] Byungchul Tak, Hyekyung Kim, Sahil Suneja, Canturk Isci, and Prabhakar Kudva. 2018. Security Analysis of Container Images Using Cloud Analytics Framework. In *Web Services -- ICWS 2018*, Springer International Publishing, Cham, 116–133.

[120] Boxiong Tan, Hui Ma, and Yi Mei. 2018. A Genetic Programming Hyper-heuristic Approach for Online Resource Allocation in Container-Based Clouds. In *AI 2018: Advances in Artificial Intelligence*, Springer International Publishing, Cham, 146–152.

[121] Boxiong Tan, Hui Ma, and Yi Mei. 2019. Novel genetic algorithm with dual chromosome representation for resource allocation in container-based clouds. In *2019 IEEE 12th International Conference on Cloud Computing (CLOUD)*, 452–456.

[122] Boxiong Tan, Hui Ma, and Yi Mei. 2019. A Hybrid Genetic Programming Hyper-Heuristic Approach for Online Two-level Resource Allocation in Container-based Clouds. In *2019 IEEE Congress on Evolutionary Computation (CEC)*, 2681–2688. DOI:https://doi.org/10.1109/CEC.2019.8790220

[123] Boxiong Tan, Hui Ma, and Yi Mei. 2020. A NSGA-II-based Approach for Multi-objective Micro-service Allocation in Container-based Clouds. In *2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID)*, 282–289. DOI:https://doi.org/10.1109/CCGrid49817.2020.00-65

[124] Boxiong Tan, Hui Ma, Yi Mei, and Mengjie Zhang. 2020. A Cooperative Coevolution Genetic Programming Hyper-Heuristic Approach for On-line Resource Allocation in Container-based Clouds. *IEEE Transactions on Cloud Computing* (2020), 1. DOI:https://doi.org/10.1109/TCC.2020.3026338

[125] Ye Tao, Xiaodong Wang, Xiaowei Xu, and Yinong Chen. 2017. Dynamic Resource Allocation Algorithm for Container-Based Service Computing. In *2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS)*, 61–67. DOI:https://doi.org/10.1109/ISADS.2017.20

[126] Jörg Thalheim, Pramod Bhatotia, Pedro Fonseca, and Baris Kasikci. 2018. Cntr: Lightweight Containers. In *2018 USENIX Annual Technical Conference (USENIX ATC 18)*, 199–212.

[127] Aparna Tomar, Diksha Jeena, Preeti Mishra, and Rahul Bisht. 2020. Docker Security: A Threat Model, Attack Taxonomy and Real-Time Attack Scenario of DoS. In *2020 10th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, 150–155. DOI:https://doi.org/10.1109/Confluence47617.2020.9058115

[128] Olufogorehan Tunde-Onadele, Jingzhu He, Ting Dai, and Xiaohui Gu. 2019. A Study on Container Vulnerability Exploit Detection. In *2019 IEEE International Conference on Cloud Engineering (IC2E)*, 121–127. DOI:https://doi.org/10.1109/IC2E.2019.00026

[129] Xili Wan, Xinjie Guan, Tianjing Wang, Guangwei Bai, and Baek-Yong Choi. 2018. Application deployment using Microservice and Docker containers: Framework and optimization. *Journal of Network and Computer Applications* 119, (2018), 97–109. DOI:https://doi.org/https://doi.org/10.1016/j.jnca.2018.07.003

[130] Bo Wang, Ying Song, Xiao Cui, and Jie Cao. 2017. Performance comparison between hypervisor- and container-based virtualizations for cloud users. In *2017 4th International Conference on Systems and Informatics (ICSAI)*, 684–689. DOI:https://doi.org/10.1109/ICSAI.2017.8248375

[131] Yulong Wang, Qixu Wang, Xingshu Chen, Dajiang Chen, Xiaojie Fang, Mingyong Yin, and Ning Zhang. 2022. ContainerGuard: A Real-Time Attack Detection System in Container-Based Big Data Platform. *IEEE Trans Industr Inform* 18, 5 (2022), 3327–3336. DOI:https://doi.org/10.1109/TII.2020.3047416

[132] Yulong Wang, Qixu Wang, Xue Qin, Xingshu Chen, Bangzhou Xin, and Run Yang. 2022. DockerWatch: a two-phase hybrid detection of malware using various static features in container cloud. *Soft comput* (2022), 1–17.

[133] Ye Wu and Haopeng Chen. 2017. Abp scheduler: Speeding up service spread in docker swarm. In *2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*, 691–698.

[134] Yiwen Wu. 2020. Exploring the relationship between dockerfile quality and project characteristics. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Companion Proceedings*, 128–130.

[135] Yiwen Wu, Yang Zhang, Tao Wang, and Huaimin Wang. 2020. Characterizing the Occurrence of Dockerfile Smells in Open-Source Software: An Empirical Study. *IEEE Access* 8, (2020), 34127–34139. DOI:https://doi.org/10.1109/ACCESS.2020.2973750

[136] Xiaojing XIE and Shyam S Govardhan. 2020. A Service Mesh-Based Load Balancing and Task Scheduling System for Deep Learning Applications. In *2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID)*, 843–849. DOI:https://doi.org/10.1109/CCGrid49817.2020.00009

[137] Jiwei Xu, Yuewen Wu, Zhigang Lu, and Tao Wang. 2019. Dockerfile tf smell detection based on dynamic and static analysis methods. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, 185–190.

[138] Kejiang Ye, Yanmin Kou, Chengzhi Lu, Yang Wang, and Cheng-Zhong Xu. 2018. Modeling Application Performance in Docker Containers Using Machine Learning Techniques. In *2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)*, 1–6. DOI:https://doi.org/10.1109/PADSW.2018.8644581

[139] Kang Yin, Wei Chen, Jiahong Zhou, Guoquan Wu, and Jun Wei. 2018. STAR: a specialized tagging approach for Docker repositories. In *2018 25th Asia-Pacific Software Engineering Conference (APSEC)*, 426–435.

Just Accepted

[140] Kang Yin, Jiahong Zhou, Wei Chen, Guoquan Wu, Jiaxin Zhu, and Jun Wei. 2018. D-Tagger: A tag recommendation approach for Docker repositories. In *Proceedings of the Tenth Asia-Pacific Symposium on Internetware*, 1–10.

[141] Takeshi Yoshimura, Rina Nakazawa, and Tatsuhiro Chiba. 2020. ImageJockey: A Framework for Container Performance Engineering. In *2020 IEEE 13th International Conference on Cloud Computing (CLOUD)*, 238–247. DOI:https://doi.org/10.1109/CLOUD49709.2020.00043

[142] Ahmed Zerouali, Valerio Cosentino, Tom Mens, Gregorio Robles, and Jesus M Gonzalez-Barahona. 2019. On the Impact of Outdated and Vulnerable Javascript Packages in Docker Images. In *2019 IEEE 26th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, 619–623. DOI:https://doi.org/10.1109/SANER.2019.8667984

[143] Ahmed Zerouali, Valerio Cosentino, Gregorio Robles, Jesus M Gonzalez-Barahona, and Tom Mens. 2019. ConPan: A Tool to Analyze Packages in Software Containers. In *2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR)*, 592–596. DOI:https://doi.org/10.1109/MSR.2019.00089

[144] Ahmed Zerouali, Tom Mens, and Coen de Roover. 2021. On the usage of JavaScript, Python and Ruby packages in Docker Hub images. *Sci Comput Program* 207, (2021), 102653. DOI:https://doi.org/https://doi.org/10.1016/j.scico.2021.102653

[145] Dong Zhang, Bing-Heng Yan, Zhen Feng, Chi Zhang, and Yu-Xin Wang. 2017. Container oriented job scheduling using linear programming model. In *2017 3rd International Conference on Information Management (ICIM)*, 174–180. DOI:https://doi.org/10.1109/INFOMAN.2017.7950370

[146] Lu Zhang, Reginald Cushing, Cees de Laat, and Paola Grosso. 2021. A real-time intrusion detection system based on OC-SVM for containerized applications. In *2021 IEEE 24th International Conference on Computational Science and Engineering (CSE)*, 138–145. DOI:https://doi.org/10.1109/CSE53436.2021.00029

[147] Qi Zhang, Ling Liu, Calton Pu, Qiwei Dou, Liren Wu, and Wei Zhou. 2018. A Comparative Study of Containers and Virtual Machines in Big Data Environment. In *2018 IEEE 11th International Conference on Cloud Computing (CLOUD)*, 178–185. DOI:https://doi.org/10.1109/CLOUD.2018.00030

[148] Rong Zhang, Yaxing Chen, Bo Dong, Feng Tian, and Qinghua Zheng. 2019. A Genetic Algorithm-Based Energy-Efficient Container Placement Strategy in CaaS. *IEEE Access* 7, (2019), 121360–121373. DOI:https://doi.org/10.1109/ACCESS.2019.2937553

[149] Rong Zhang, A-min Zhong, Bo Dong, Feng Tian, and Rui Li. 2018. Container-VM-PM Architecture: A Novel Architecture for Docker Container Placement. In *Cloud Computing -- CLOUD 2018*, Springer International Publishing, Cham, 128–140.

[150] Yang Zhang, Gang Yin, Tao Wang, Yue Yu, and Huaimin Wang. 2018. An insight into the impact of dockerfile evolutionary trajectories on quality and latency. In *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, 138–143.

[151] Yan Zheng, Weiyu Dong, and Jiangtao Zhao. 2021. ZeroDVS: Trace-Ability and Security Detection of Container Image Based on Inheritance Graph. In *2021 IEEE 5th International Conference on Cryptography, Security and Privacy (CSP)*, 186–192. DOI:https://doi.org/10.1109/CSP51677.2021.9357496

[152] Jiahong Zhou, Wei Chen, Guoquan Wu, and Jun Wei. 2019. SemiTagRec: a semi-supervised learning based tag recommendation approach for Docker repositories. In *International Conference on Software and Systems Reuse*, 132–148.

[153] Ruiting Zhou, Zongpeng Li, and Chuan Wu. 2018. Scheduling Frameworks for Cloud Container Services. *IEEE/ACM Transactions on Networking* 26, 1 (2018), 436–450. DOI:https://doi.org/10.1109/TNET.2017.2781200

[154] Documentation · Wiki · AppArmor / apparmor · GitLab. Retrieved April 12, 2022 from https://gitlab.com/apparmor/apparmor/-/wikis/Documentation

[155] Container Vulnerability Scanning • Anchore. Retrieved April 12, 2022 from https://anchore.com/container-vulnerability-scanning

[156] MicroScanner: New Free Image Vulnerability Scanner for Developers - Aqua. Retrieved April 12, 2022 from https://www.aquasec.com/news/microscanner-new-free-image-vulnerability-scanner-for-developers/

[157] Docker Image/Container Security Scan with Clair — Installation | by Kinjal Rathod | System Weakness. Retrieved April 12, 2022 from https://systemweakness.com/docker-image-container-security-scan-with-clair-installation-355f80201ef5

[158] Cilium - Linux Native, API-Aware Networking and Security for Containers. Retrieved April 12, 2022 from https://cilium.io/

[159] y-cruncher - A Multi-Threaded Pi Program. Retrieved April 13, 2022 from http://www.numberworld.org/y-cruncher/

[160] NBench - Wikipedia. Retrieved April 13, 2022 from https://en.wikipedia.org/wiki/NBench

[161] LINPACK. Retrieved April 13, 2022 from http://www.netlib.org/linpack/

[162] The noploop CPU Benchmark. Retrieved April 13, 2022 from https://www.brendangregg.com/blog/2014-04-26/the-noploop-cpu-benchmark.html

[163] Parallel BZIP2 (PBZIP2). Retrieved April 13, 2022 from http://compression.ca/pbzip2/

[164] Geekbench 5 - Cross-Platform Benchmark. Retrieved April 13, 2022 from https://www.geekbench.com/

[165] Sysbench - Gentoo Wiki. Retrieved April 13, 2022 from https://wiki.gentoo.org/wiki/Sysbench#Using_the_CPU_workload

[166] SPEC CPU® 2006. Retrieved April 13, 2022 from https://www.spec.org/cpu2006/

[167] STREAM Benchmark - AMD. Retrieved April 13, 2022 from https://developer.amd.com/spack/stream-benchmark/

[168] GitHub - cruvolo/ramspeed-smp: RAMspeed/SMP, a cache and memory benchmarking tool. Retrieved April 13, 2022 from https://github.com/cruvolo/ramspeed-smp

[169] Iozone Filesystem Benchmark. Retrieved April 13, 2022 from https://www.iozone.org/

[170] dbench. Retrieved April 13, 2022 from https://dbench.samba.org/doc/dbench.1.html

Just Accepted

[171] Bonnie++ - Wikipedia. Retrieved April 13, 2022 from https://en.wikipedia.org/wiki/Bonnie%2B%2B

[172] 1. fio - Flexible I/O tester rev. 3.27 — fio 3.27-169-g1953e1-dirty documentation. Retrieved April 13, 2022 from https://fio.readthedocs.io/en/latest/fio_doc.html

[173] netpipe_web. Retrieved April 13, 2022 from https://netpipe.cs.ksu.edu/

[174] iPerf - The TCP, UDP and SCTP network bandwidth measurement tool. Retrieved April 13, 2022 from https://iperf.fr/

[175] netperf - networking benchmark software - LinuxLinks. Retrieved April 13, 2022 from https://www.linuxlinks.com/netperf/

[176] Container Monitoring Tools | Sematext. Retrieved April 13, 2022 from https://sematext.com/container-monitoring/

[177] docker stats | Docker Documentation. Retrieved April 13, 2022 from https://docs.docker.com/engine/reference/commandline/stats/

[178] Sensu Plugins. Retrieved April 13, 2022 from https://sensu-plugins.io/

[179] Cloud Native & Secure DevOps Platform | Sysdig. Retrieved April 13, 2022 from https://sysdig.com/platform-architecture/

[180] Overview | Prometheus. Retrieved April 13, 2022 from https://prometheus.io/docs/introduction/overview/

[181] Monitoring Docker container metrics using cAdvisor | Prometheus. Retrieved April 13, 2022 from https://prometheus.io/docs/guides/cadvisor/

[182] Container Monitoring | Datadog. Retrieved April 13, 2022 from https://www.datadoghq.com/product/container-monitoring/

[183] Container Monitoring: Essential Tools + Best Practices | Scout APM Blog. Retrieved April 13, 2022 from https://scoutapm.com/blog/container-monitoring

[184] Container monitoring | Dynatrace. Retrieved April 13, 2022 from https://www.dynatrace.com/solutions/container-monitoring/

[185] Container Monitoring Tool – Monitor Docker Containers | AppOptics. Retrieved April 13, 2022 from https://www.appoptics.com/use-cases/container-monitoring

[186] Welcome Page. Retrieved April 13, 2022 from https://www.nuttcp.net/Welcome%20Page.html

Just Accepted

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-18119-w

A systematic literature survey on skin disease detection and classification using machine learning and deep learning

Rashmi Yadav[1] · **Aruna Bhat**[1]

Received: 15 June 2023 / Revised: 17 September 2023 / Accepted: 1 January 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
The world population is growing very fast and the lifestyle of human beings is changing with time and place. So, there is a need for disease management which includes disease diagnosis, its detection and classification, cure and lastly for future disease prevention. The outermost protective layer of a human body is the skin. Skin not only impacts a person's health but also psychologically impacts one's life. Computer-aided systems are very helpful in skin disease detection and classification and their application is growing rapidly in healthcare. This literature review paper aims to help the researchers to get a synthesized and appropriate information for the same. We have included papers from 2021 to 2023 for the review from the Scopus database. 45 studies are selected for the review of which 32 studies use deep learning techniques, 11 use machine learning techniques and 2 studies use a hybrid approach. The studies are compared on various parameters like models, datasets, and performance metrics. The work also identified some of the challenges like dealing with noise and also explained disease symptoms.

Keywords Skin diseases · CNN · Deep learning · Systematic literature · Machine learning

1 Introduction

Skin diseases are among the general medical conditions, affecting millions of people worldwide. For efficient treatment and the avoidance of consequences, skin illnesses must be identified early and correctly diagnosed. Skin diseases are a prevalent health issue caused by numerous factors such as fungal infections, bacterial or viral infections, and allergic reactions, these may transform various changes in texture, color of the skin, and if left untreated, they may develop into more severe conditions such as skin cancer. As skin diseases are chronic and infectious, it is crucial to diagnose them early to prevent further spread and development. However, diagnosing and treating skin diseases can be

✉ Aruna Bhat
aruna.bhat@dtu.ac.in

Rashmi Yadav
rashmiyadav.1512@gmail.com

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

Published online: 26 February 2024

time- consuming and expensive, resulting in both physical and financial costs for patients [7]. Melanoma and non-melanoma are two subtypes of skin cancer. This is a deadly, hazardous and rare type of skin cancer that starts in melanocyte cells. Melanoma typically develops in exposed areas like the face, neck, hands, lips, and other places that are exposed to UV radiation from the sun. The only prevention for this skin cancer is if it can be detected at a premature phase. Nodular melanoma, Superficial spreading melanoma, Acral lentiginous and Lentigo malignant are some rare and deadly forms of skin cancer [13] whereas basal cell carcinoma, squamous cell carcinoma and sebaceous gland carcinoma are some common non-melanoma skin cancers. The biopsy is one of the methodologies practised for the treatment of skin cancer, in this a sample of the affected skin lesion is removed and examined medically to see if it is cancerous. This procedure is challenging, drawn out, time- consuming and lengthy. Computer-based technology has the ability to give cost-effective, comfortable, safe, painless and speedy diagnoses of skin cancer or other disease symptoms. However, skin disease detection and classification can be challenging because of the wide range of possible diseases, the high variability in appearance and severity, and the limited availability of expert dermatologists. For skin disease detection and its classification machine learning and deep learning techniques have emerged as promising tools, leveraging large datasets of annotated images and sophisticated models. Generally, in the above-stated techniques we train the model, extract features of the image then predicted the outputs. In this process feature extraction is an important step, as outcomes depend on it. In [110] proposes product attribute value extraction names MixPave (Mixed Prompts for few shots Attribute Value Extraction), which is basically introduced for small trainable parameters, and also helps in reducing the overfitting problem. So, it may help with skin detection and classification problem.

1.1 Skin disease diagnosis

The doctor used the conventional approach to diagnose the patient's skin condition based on his expertise, experience, and understanding of the characteristics, traits and features visible in the dermatoscopic images. We can sum up the diagnostic procedure as follows: first, using histological analysis and dermoscopy after the doctor performs a visual diagnostic to gather the essential patient information. Dermoscopy is a high-definition imaging technique that is non-invasive and can be used to examine the structure of the skin where the superficial dermis and lower epidermis meet. Doctors examined the characteristics of skin lesions, their allocation, positioning, colour, diameter and border, area, shape and structure of pigmented skin lesions using dermatoscopy techniques like ABCDE rule [32], a seven-point checklist[69], chaos and Clues[69], three-point checklists [69], and CASH (colour, architecture, symmetry, and homogeneity) [69]. Because skin lesions are similar in form, texture, edge contour, colour, and other traits, and diseased tissues differ across different patients, only expert dermatologists can reliably diagnose pigmented skin disease and its pathological symptoms. The patient's requirement for medical resources is not at all satisfied by this approach of depending on expertise for diagnosis. This whole process usually takes 4-5 days, which includes sampling, diagnosis, histopathological examination then the report of the patient which in summary consumes a lot of time and also influences the patient physically and psychologically.

The disadvantages of the conventional dermatological diagnosis are as follows: First and foremost are the restrictions of medical institutions. Dermatologists who have experience with skin are scarce. Due to the imbalance between the rate of dermatologist growth and

the frequency of skin illnesses, there are currently many patients and not enough skilled dermatologists. Second, the diagnosis is less accurate. Dermatologists with expertise may diagnose the same patient in a diverse way based on their personal views and work experiences. Even the same doctor, is affected by several factors like environment, fatigue, and personal factors, which result in different diagnoses for the same dermatoscopic image. Third, due to the characteristics of the skin, there are both minor gaps between categories and enormous gaps within them in dermatoscopic images. As a result, the diagnosis has a higher chance of being incorrect or missing a diagnosis, which lowers its accuracy.

1.2 Image recognition of skin lesion using Machine learning (ML) and Deep Learning (DL)

For skin disease detection and classification, ML and DL can be powerful tools. These tools make it possible to analyse large data and find patterns that might not be immediately obvious to the naked eye. Some of the ML and DL techniques which show high-quality results such as Random Forest(RF) [61], Bernoulli Na¨ıve Bayesian (BNB) [101], Knowledge Distillation [105], Residual Network (ResNet)[54], DenseNet201[43], Grasshopper Optimization[98], EfficientNet[92] and many more are listed in Table 3.

Skin, hair, and nail issues are referred to as skin illnesses or dermatological disorders. The author utilises [108] five convolutional neural network (CNN) models for facial skin classification: ResNet50, Inception V3, DenseNet 121, Xception, and Inception-ResNet V2. Skin illnesses have also had an impact on the patient's psychology. Numerous factors, such as family history, surroundings, and underlying medical issues, might contribute to the development of these illnesses. Skin conditions can affect any portion of the body and have a number of causes. While some skin diseases are chronic and require ongoing management, many can be treated effectively with medication and lifestyle changes. In order to avoid long-term consequences, it's crucial to consult a dermatologist as soon as you suspect you have a skin condition. There are mainly two types of skin disorders: acute and chronic. Acute disorders like Acne, contact dermatitis, cold sore, measles etc. are shown in Figure 1 and which are chronic including mainly skin tumours and their types.

With the help of computer-aided diagnosis, there are two approaches, one is ML and another is DL. As DL is the subset of ML. A model is trained on a massive annotated dataset of labelled images, enabling it to recognise patterns and traits associated with various skin conditions. After training, the model may be used to categorise new images and offer a diagnostic based on the patterns it has discovered. This process can be done by some popular and publicly available databases such as ISIC achieves[56], PH2[1], ISBI achieves[90] and others are listed in Table 4. Another approach is to use DL algorithms to analyze medical records and other patient data. This can be particularly useful in cases where there is not enough data available to train a machine-learning model on images alone. By examining a range of patient data, such as symptoms, medical history, and other clinical information, deep-learning algorithms can identify individuals who are at risk for particular skin illnesses and suggest suitable treatment options. For the classification and identification of skin illnesses, several publications adopt a single strategy, such as, [75] which pre-processes data with a custom CNN and an enhanced super-resolution generative adversarial network (ESRGAN) to improve accuracy, another author [14] used EfficientNet-B6 which results in high accuracy on HAM100000 database and ISIC achieves respectively. Some authors employ a multistage [73] framework to categorise skin lesions; in the first stage, they used a contrast enhancement-based bio-inspired multiple exposure fusion approach to

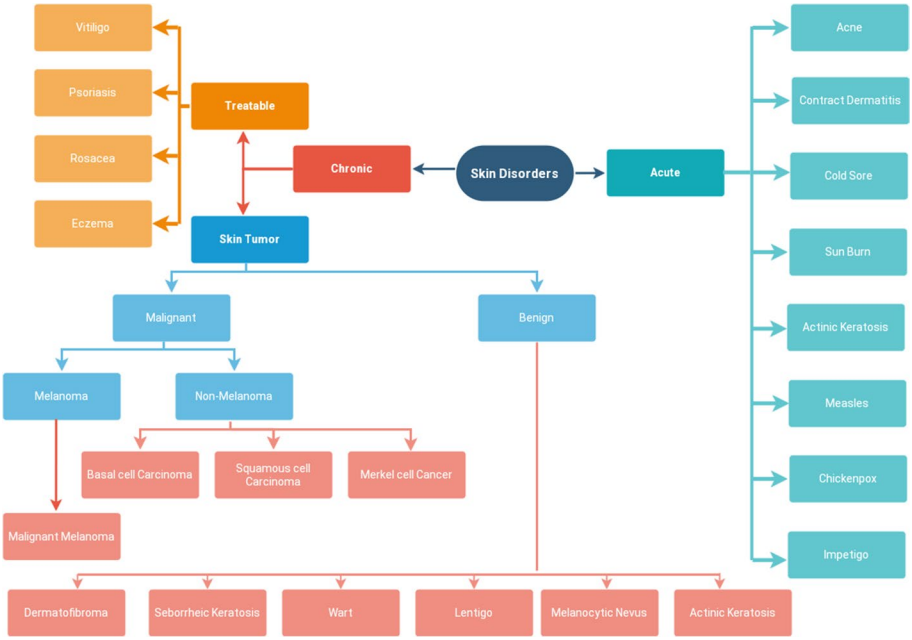


Fig. 1 Skin taxonomy (Sources: Authors)

pre-process the input image. A 26-layered CNN is employed in the second stage, and four previously trained CNN are tweaked and trained in the third stage. The deep feature vectors are taken from CNN models and fused CNN models in the fourth stage, and at the final stage, Poisson distribution is used for feature selection which results in 98.98% accuracy.

Traditional machine learning algorithms like RF [74] shows high-quality results on real data sets also. The authors collected data from a popular private hospital in Chennai, where an average of 2000 to 6000 patients visits every day. The data collected from the hospital are of different formats like audio, video, image and structure. At pre-processing step, they converted data in machine-readable form i.e., in binary language. Next missing value dealing so that data shouldn't be lost and rows/columns can't be eliminated otherwise it would lower down the sensitivity of the data. The process of feature selection is primarily intended to increase the model's efficiency and cost-effectiveness. So, lastly, for classification, K-Nearest Neighbour (KNN), Support Vector Machine (SVM) and RF are considered. RF shows high results with 97% accuracy and has a new feature set compared to other models.

SVM with RF is used by Murugan et al. [76] with 89.31% accuracy. The authors' method involves image processing. Mean shift segmentation and a median filter are utilised for the pre-processing. Moment invariant features, gray level co-occurrence matrix (GLCM) features, and grey level run matrix are used in the feature selection approach. Utilising SVM, RF combined SVM, and probabilistic neural network, the retrieved features are categorised. The ISIC dataset is used to evaluate the performance.

Hatem [47] uses KNN for stimulation, the platform used is MATLAB. The processes involved in the detection of skin disorders are initial processing, splitting, extraction of features and categorization. The primary goals of pre-processing, which include hair removal and picture enhancement, are to enhance and increase the quality of an image. The

threshold approach for segmenting images includes setting the intensity of each pixel. The feature is extracted, and classification is done by KNN. The author also shows results in GUI and achieved with 98% accuracy.

Optimization algorithms are also highly efficient and focus on finding solutions to complex problems. The class of metaheuristic algorithms known as" nature-inspired optimisation algorithms" is inspired by natural phenomena and how they behave to classify, identify, and segment skin diseases. The sparrow search algorithm [13] is one method the author uses to optimise the hyperparameters of eight CNN models that have been previously trained using five open-source datasets. The experiment is split into segmentation and classification, with MobileNet's pre-trained model achieving the greatest accuracy (98.83%) on the classification challenge using the HAM10000 melanoma dataset. Also, the proposed approach is compared to 13 related studies and shows a competitive state-of-art.

For the identification of malignant melanoma, a suggested fractal neural network-based galactic swarm optimisation (FNN- GSO) [62] algorithm has been developed. The primary components of the proposed study are the pre-processing, segmentation, feature extraction, and classification of melanoma. In pre-processing stage removal of noise and contrast enhancement are accomplished. The adaptive watershed algorithm is used for segmentation. Image features are extracted using the DWT- GLCM model and FNN-GSO for malignant melanoma classification. Implementation work is done on MATLAB software with PH2, HAM10000 and ISIC dataset with 89.06% Jaccard index, dice coefficient of 90.56% and Matthew coefficient of 90.76%.

Adla et.al. [3] introduced DL with a class attention layer based on Computer-aided diagnosis (CAD) for skin lesion detection and classification (DLCAL-SLDC). Dull razor approach and average median filtering are used for hair removal and noise removal respectively for the pre-processing step. Using a Capsule network and Adagrad optimizer, features are retrieved from the segmented images. Finally, classification is achieved using the SSO-CSAE model, a convolutional sparse autoencoder (CSAE) based on Swallow Swarm Optimisation (SSO), which achieves 99.1% specificity, 94.5% sensitivity, and 98.50% accuracy on the ISIC 2017 dataset.

Usma et al. [100] suggested a hybrid and optimal feature selection-based improved automated system. The whole experiment is done on HAM10000. Four phases were included in the strategy; the first stage involved balancing the dataset by using three alternative transformation approaches. In the second stage, DarkNet53 and InceptionV3 are retrained using transfer learning. Deep features are retrieved from the dataset in the third stage, reduced using moth flame optimisation (MFO), and classification is carried out in the fourth stage using cubic SVM, quadratic SVM, linear SVM, KNN, and ensemble subspace discriminant. 95.9% is the highest accuracy gained on cubic SVM. They compare their techniques without MFO and other state-of-art approaches.

Kumar et. al. [67] proposes an optimization-based skin cancer detection using dermoscopic images. Initial pre-processing steps include the region of interest extraction and the use of a Gaussian filter, followed by segmentation using a U-RP net model. SqueezeNet, which was trained using the suggested Aquila whale optimisation (AWO), is then used to extract the features. This method achieves 92.5% accuracy, 92.1% sensitivity, and 91.7% specificity. The Skin Cancer: Malignant vs. Benign dataset from ISIC 2021 is used for the simulation study.

A non-parametric approach proposed by [104], is deep nearest centroids (DNC) in which authors compared it to a parametric approach giving 5 main merits i.e. simplicity and transparency compared to deep neural networks, automated discovery of underlying data structure, direct supervision of representation learning, better transferability

and ad-hoc explainability. This approach improves the efficacy and interpretability of the experiment.

2 Literature Survey

The Systematic literature review (SLR) basically consists of the identification, evaluation (screening) and interpretation of a particular topic area. To write this SLR on Skin Diseases (SD) we have followed a sequence of steps depending on different approaches using PRISMA 2020 guidelines shown in Figure 2. The first step needed is to identify the research questions so that we can get the necessary keywords to search related recent literature to our SLR. Then we collected the different literature sources from articles and this step is called searching. We have collected the research studies from different sources named IEEE, SpringerLink, Elsevier and MDPI and every paper's sources are described in Table 1. The collected research studies have journal papers only. Next, we have defined the Inclusion criteria and exclusion criteria to exclude the research studies. In the next section,

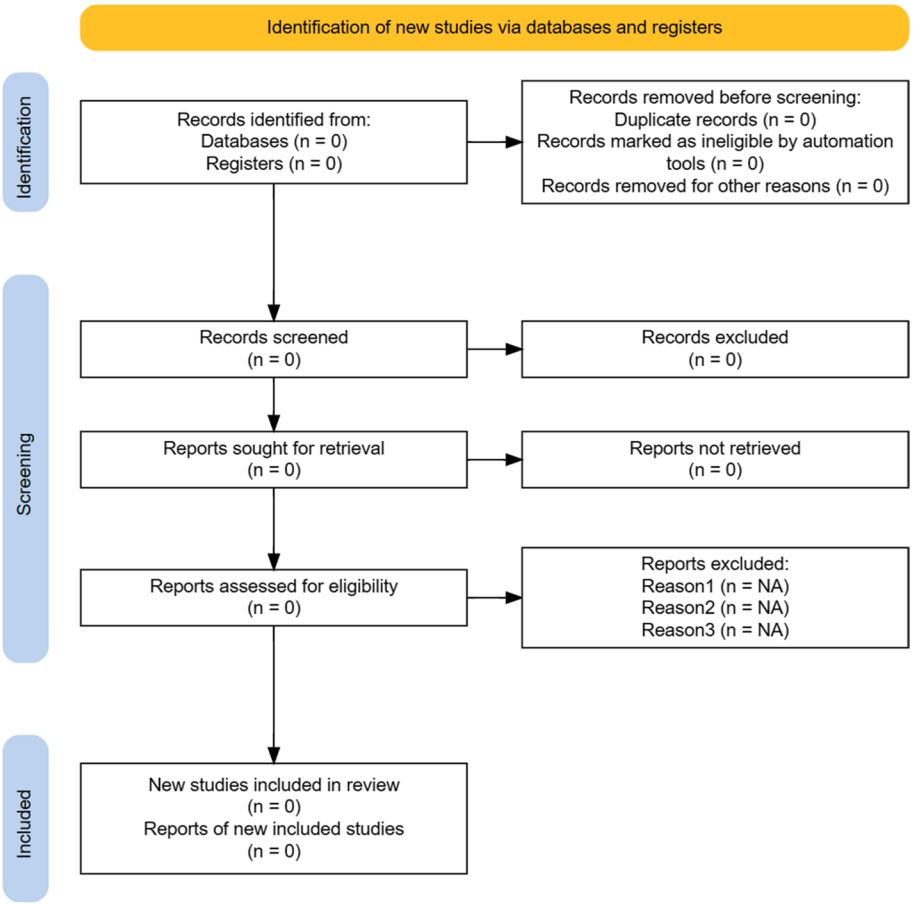


Fig. 2 PRISMA flowchart

Table 1 Journal/conference information

Research Study	Reference	Year of Publication	Publisher	Type
RS1	[102]	2023	Elsevier	Journal
RS2	[73]	2023	Elsevier	Journal
RS3	[11]	2023	Elsevier	Journal
RS4	[6]	2023	Elsevier	Journal
RS5	[105]	2023	Elsevier	Journal
RS6	[42]	2023	Elsevier	Journal
RS7	[50]	2023	Springer	Journal
RS8	[84]	2023	MDPI	Journal
RS9	[75]	2023	MDPI	Journal
RS10	[14]	2023	MDPI	Journal
RS11	[93]	2023	MDPI	Journal
RS12	[40]	2023	Elsevier	Journal
RS13	[54]	2023	Elsevier	Journal
RS14	[13]	2023	Springer	Journal
RS15	[37]	2023	Springer	Journal
RS16	[91]	2023	Springer	Journal
RS17	[72]	2023	Springer	Journal
RS18	[9]	2022	MDPI	Journal
RS19	[8]	2022	MDPI	Journal
RS20	[85]	2022	MDPI	Journal
RS21	[95]	2022	Elsevier	Journal
RS22	[62]	2022	Springer	Journal
RS23	[47]	2022	Springer	Journal
RS24	[3]	2022	Springer	Journal
RS25	[52]	2022	Springer	Journal
RS26	[21]	2022	Elsevier	Journal
RS27	[100]	2022	MDPI	Journal
RS28	[106]	2022	IEEE	Journal
RS29	[10]	2022	Springer	Journal
RS30	[79]	2022	MDPI	Journal
RS31	[17]	2022	MDPI	Journal
Rs32	[67]	2022	Elsevier	Journal
RS33	[109]	2022	Elsevier	Journal
RS34	[74]	2022	Springer	Journal
RS35	[71]	2022	Elsevier	Journal
RS36	[86]	2022	Springer	Journal
RS37	[78]	2022	Elsevier	Journal
RS38	[87]	2022	Springer	Journal
RS39	[77]	2022	MDPI	Journal
RS40	[90]	2022	Springer	Journal
RS41	[41]	2022	IEEE	Journal
RS42	[55]	2021	Elsevier	Journal
RS43	[65]	2021	Elsevier	Journal
RS44	[76]	2021	Elsevier	Journal
RS45	[64]	2021	Elsevier	Journal

we have defined the research questions as we have discussed above. These research questions are very useful to collect the studies and summarize the review.

2.1 Research Questions

This section consists of the identification of research questions which are defined on the basis of the summarization of the topic so that keywords can be generated and which will be useful for the search of the studies on different platforms. We find out a total of 6 research questions which are defined below on the basis of which the whole SLR will move forward to tell you about skin disease detection and classification. The research questions are given below:

RQ1. What are the available literature sources for skin disease detection and classification?
RQ2. What are the different skin diseases present?
RQ3. What is the most recent development in deep learning and machine learning architectures?
RQ4. Which data sets are being utilized to train the models?
RQ5. What are the main performance measures used for validating skin-related diseases?
RQ6. How can we deal with noisy data coming from different sources?
RQ7. What can be the future research directions for this field?

2.2 Study Selection

With the aid of the PRISMA tool, which is depicted in Figure 2, the things that were most important to the objective of this systematic literature review were able to be identified through the selection procedure. Our study found that a similar article was only looked at once if it appeared in more than one domain. To make sure that the findings of the current systematic literature review produced a clear and impartial conclusion. For the current systematic literature review's findings to be accurate and led to conclusions that were impartial, understandable and clear, the substance of the publications selected for the final study was evaluated. We finished the assessment to determine if we would ultimately be included or excluded. Individual assessment inconsistencies were then discussed and rectified. After finding the articles, the first step was to get rid of any duplicates and titles that weren't relevant to the subject. One or more of the exclusion criteria (EC) that were met by the study were disqualified, and the search for String was the only study that met the inclusion criteria (IC).

2.2.1 Inclusion Criteria (IC)

1. Research articles match the search queries.
2. Studies which have open access or can be accessed by an institutional proxy (account or network).
3. The study span is from 2021 to March 2023.
4. Only journals are studied.
5. Final articles are reviewed.
6. The digital database contains the entire text.

7. The aim is relevant to Skin disease detection and classification using ML or DL.

2.2.2 Exclusion criteria (EC)

1. We avoid using identical research publications.
2. Literature reviews, periodicals, conferences, monographs, reports, surveys, book chapters, monographs, thesis and interview-based papers are excluded.
3. Research articles which are published before 2021.

The total acquired research papers and studies are found on the above-introduced standard for inclusion and exclusion. Figure 3 is a PRISMA flow chart created using the PRISMA model. Figure 3 is useful to provide the idea of the study selection procedure.

We have answered all the research questions defined by us in the above section. Data from selected 45 studies have been extracted and synthesized. The responses to the research question are provided in accordance with the data in the next section, and outcomes after a summary of all papers has been generated as an Excel sheet. For the reference of each paper, we used RSi so that we don't need to use the whole paper again and again. The reference for each paper is given in Table 1.

3 Detailed Discussion of Literature Survey

In order to identify and categorise skin diseases, we have chosen 45 search articles. Figure 4 depicts the useful keywords to find out the research studies and these keywords are also available in the papers. After that, we have summarised the study based on their literature sources and year as shown in figure

5, types of skin diseases, model/techniques/architecture used for it, the dataset used in it and also we have mentioned some publicly available dataset which any of the selected studies is used, on the basis of their performances and lastly we discussed some techniques to deal with noise.

3.1 Research Questions

3.1.1 RQ1: What are the available literature sources for skin disease detection and classification?

The major number of papers is from the year 2022. All the research articles are from journals only. A maximum number of journals from Elsevier followed by Springer. The number of year-wise and publication-wise studies are shown in Figures 5 and 6. The literature source, year and type of article are summarized in Table 1.

3.1.2 RQ2: What are the different skin diseases present?

In Figures 1 and 7 we have shown all the skin diseases which we have discussed in Table 2. Skin cancer can be divided into two subtypes: melanoma and non-melanoma. According to Table 2 majority of research is towards melanoma detection and classification, as it is a rare, deadly and dangerous disease which can only be cured at the initial stage. The main reason for melanoma is UV exposure.

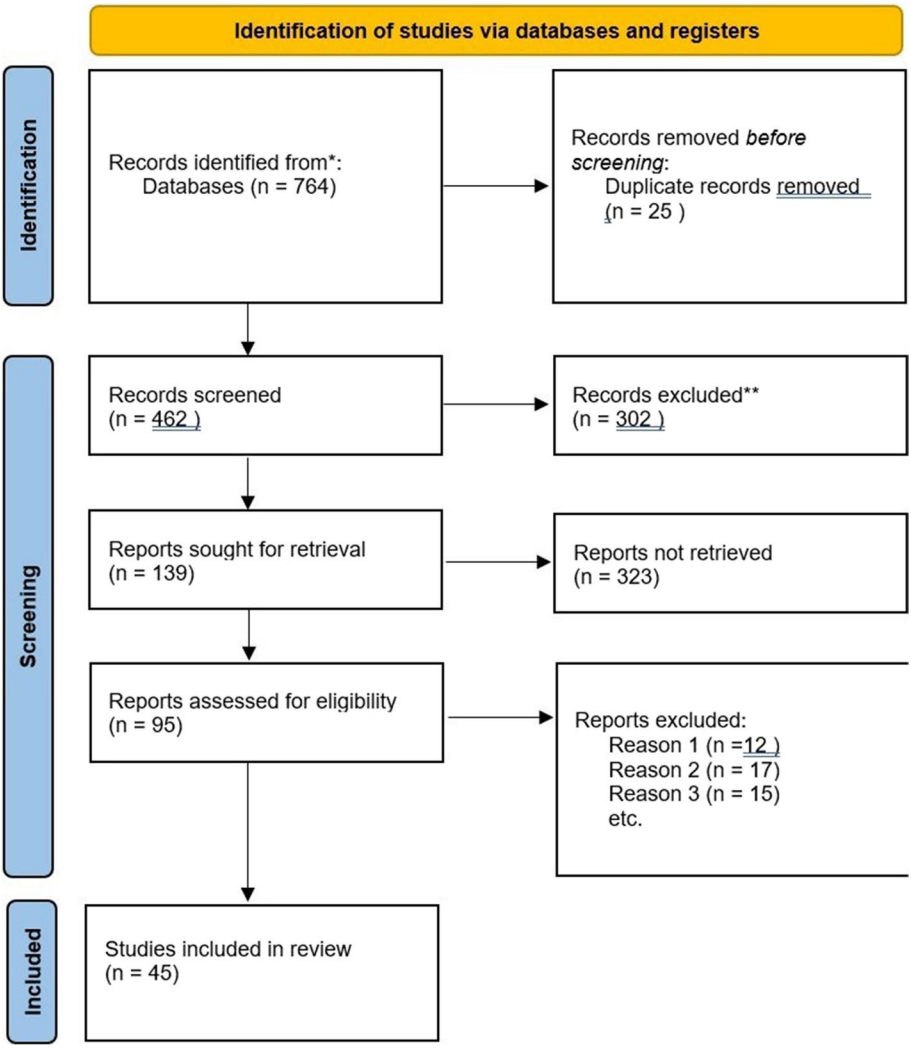


Fig. 3 Selection of studies based on PRISMA flow chart

3.1.3 RQ3: What is the most recent development in deep learning and machine learning architectures?

There are a lot of techniques which trending in ML and DL. Some of the common and explained in this SLR are as follow:

1. **Convolutional Neural Networks (CNNs)** [68]: CNN is feed-forward neural network which is highly used in the applications of computer vision especially in image classification, object detection, self-driving cars, facial emotion recognition and many more. As depicted in Fig. 8, the fundamental architecture of CNN consists of a convolutional layer, a pooling layer, and a fully connected layer. These layers can go to 25–30 layers,

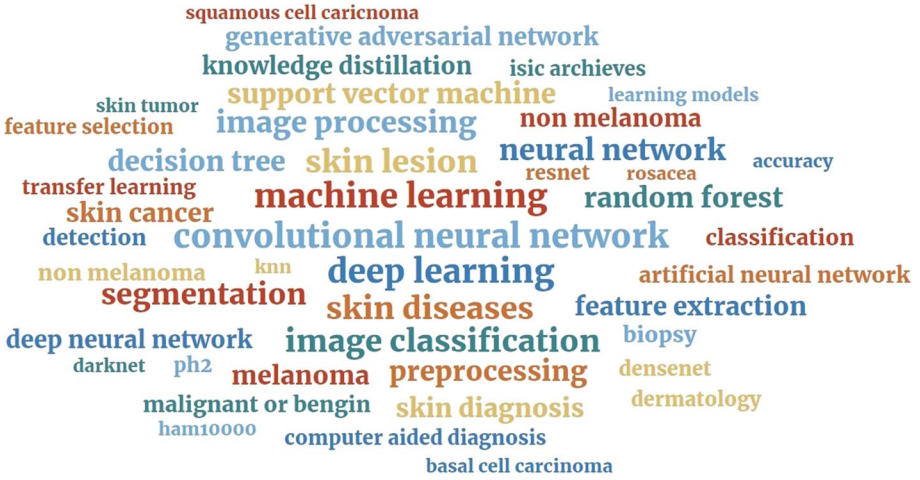


Fig. 4 Keywords present in all the research article for the SLR

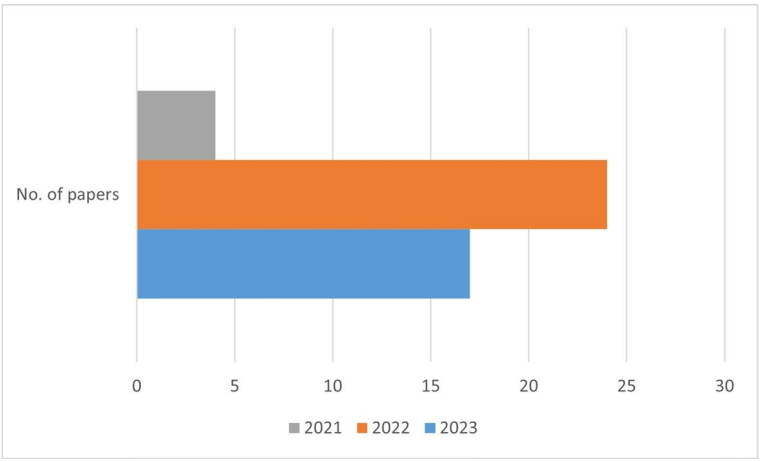


Fig. 5 Number of studies each year

these digits may vary depending on the problem which extracts features from the image. CNNs have been heavily used to categorise images of skin lesions.

2. **Support Vector Machines (SVMs)** [18]: SVM is a part of the supervised learning method, which is applied to classification problems in ML. In order to make it simple to categorise fresh data that is received in the future, SVM attempts to construct the ideal line or decision boundary that can divide n-dimensional space into categories. This optimal decision boundary is known as a hyperplane. Based on features taken from images of skin lesions, SVMs have been used to categorise various skin disorders.
3. **Random Forests (RF)** [20]: A classification decision is made using the ensemble learning technique known as Random Forests, which mixes various decision trees. They have been used to classify skin illnesses based on characteristics extracted from photographs of skin lesions.

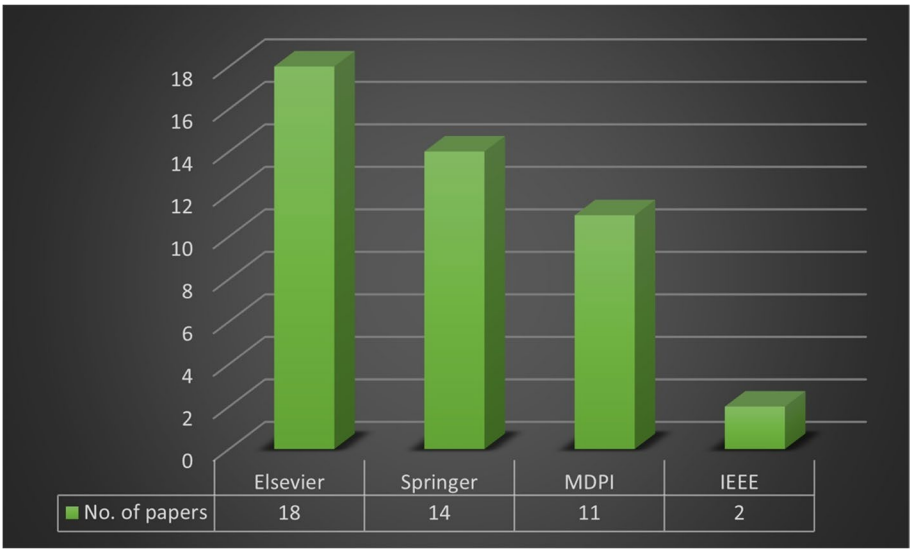


Fig. 6 Count of Literature source



Fig. 7 Samples of skin diseases described in the paper[23, 26, 27, 89]

Table 2 Different types of skin diseases

Diseases	Research Study	Symptoms of Disease on Skin
Melanom (MEL) [25]	RS1, RS2, RS3, RS4, RS5, RS6, RS7, RS9—RS20, RS22, RS24- RS45	A cancerous and life-threatening mole or spot on the skin that is asymmetrical, irregular border, different colours (shades of brown, red, black, white or blue), size is larger than 6 mm in diameter, mole shape, size, color changes over the time, a mole can be itchy, painful or bleeding
Seborrheic Kerato- sis (SK) [29]	RS1, RS21, RS24, RS37, RS44	It is a non-cancerous skin growth which can be from few mm to cm in diameter, growth may be waxy, scaly or rough in appearance, color can be light tan, brown, black, or even white or yellow in color, found in chest, back or face with no itching or pain
Melanocytic Nevi (NV) [81]	RS1- RS4, RS7, RS9, RS15, RS16, RS18—RS20, RS24, RS27 - RS30, RS35, R37, RS41—RS43, RS45	Benign (noncancerous) skin growths that can appear anywhere on the body, harmless, round or oval in shape can vary from tiny dot to larger spot, color can be brown, black, or tan with clear border and symmetric in nature
Basal Cell Carci- noma (BCC) [24]	RS1- RS6, RS9, RS15—RS20, RS22, RS24, RS26-RS30, RS33, RS35, RS37—RS39, RS42- RS45	Cancerous, slow growing, Pearly or waxy bump on the skin, Pink or flesh-colored lesion with raised borders, does not heal or bleeds easily, occurred on face, neck, and hands due to exposure in sun
Squamous Cell Carcinoma (SCC)[30]	RS5, RS11, RS17, RS22, RS25, RS26, RS31, RS32, RS33, RS42, RS45,	Cancerous, Scaly red patch or wart-like growth, does not heal or bleeds easily, commonly happens in parts of the skin like the face, ears, neck, and hands that are frequently exposed to the sun
Actinic Keratosis (AKIEC) [22]	RS1—RS6, RS9, RS15, RS16, RS18—RS20, RS25—RS32, RS35, RS37, RS43, RS45	Non-cancerous, Rough, scaly patch or bump on the skin, Pink, red, or brown in color, normally occurred on face, ears, scalp, and arms due to exposure in sun
Benign Keratosis- Like Lesions (BKL) [28]	RS1—RS5, RS7, RS9, RS15, RS18—RS20, RS27—RS30, RS35, RS37, RS39, RS42, RS43, RS45	Non-cancerous, Raised, rough, or scaly growth on the skin, tan, brown, or black in color, occurred at areas which are to exposed to the sunfaces, scalp and hands

Table 2 (continued)

Diseases	Research Study	Symptoms of Disease on Skin
Dermatofibroma (DF) [80]	RS1—RS5, RS9, RS15—RS20, RS27—RS30, RS35, RS37, RS42, RS43, RS45	Non-cancerous, Firm, raised bump on the skin, Usually brown in color, May itch or be tender to the touch, commonly occurs on the legs and arms
Vascular Lesions (VASC)[82]	RS1 RS5, RS9, RS15, RS16, RS18—RS20, RS27—RS30, RS35, RS37, RS42, RS43, RS45	Non-cancerous, red, purple, or blue discolouratio of the skin, flat or raised, can be brought on by a variety of causes, such as heredity, sun exposure, and skin injuries
Rosacea [27]	RS8	Redness on the core features of the face, including the chin, cheeks, forehead, and nose. Face lumps that are small, red, and packed with pus, visible blood vessels on the face with burning sensation, dry, irritated, or swollen eyes, thickened, bumpy skin on the nose, called rhinophyma (more common in men)
Eczema [23]	RS13, RS23	Chronic skin condition with Intense itching, especially at night, dry, cracked or scaly, or thickened skin, on the skin, red or brownish-gray spots, little- raised lumps that, if touched, may leak fluid and crust over, skin that is swollen, sensitive, or in- flamed, dark-colored patches of skin that may be more common in people with dark skin tones
Psoriasis [26]	RS13, RS23,	Chronic autoimmune disorder, dry, cracked, painful, Red, raised, inflamed patches of skin, frequently coated in scales of silver or white, Itching, burning or soreness in affected areas, ridged, pitted, or thickened nails stiff and swollen joints, small areas of scaling (often noticed in youngsters)
Others	RS14, RS21, RS23, RS24, RS34	NA

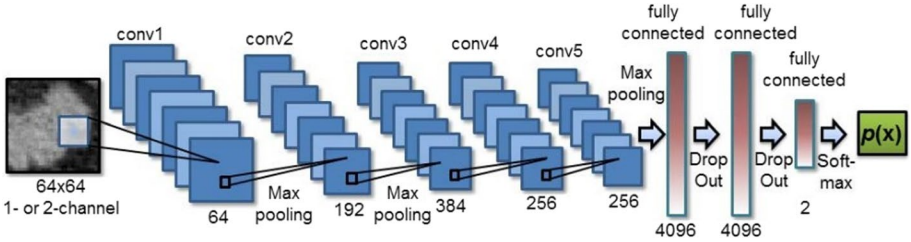


Fig. 8 Basic CNN architecture [39]

4. **Transfer Learning (TF)** [66]: Transfer learning is currently very popular in DL due to the reuse of knowledge of a pre-trained model which can be applied to a similar task on which already a model is trained. It will assist in lowering the model's computational expense and make it more efficient and gives high accuracy.
5. **Ensemble Learning**[19, 20, 88, 107]: Ensemble learning consists of bagging, boosting and stacking. It states that use multiple learning algorithms to get better results.

CNN architecture is mainly used for image classification which is part of deep learning. VGG 19[94], InceptionV3[96], AlexNet, GoogleNet, ResNet[48], DenseNet[51], EfficientNet[97], SqueezeNet[53] and DarkNet are popular CNN architectures in our paper. Every CNN classifier has its own pros and cons depending on the problem. The choice of CNN classifier depends on distinct factors like size and complexity of the problem, computational cost available, complexity and the size of the dataset, available pre-trained models, tuning of hyperparameter which includes activation function, optimizer, learning rate etc. These are some factors on which researchers can think before applying any classifier. These also impacted the accuracy of the model.

Generative Adversarial Networks (GAN) and Knowledge distillation (KD) are part of machine learning techniques, and nowadays these are emerging fastly and showing promising results in this field which are explained in this section and used in Table 3

1. **Knowledge Distillation (KD)** [49] Fig. 9 illustrates the usage of KD, an ML method, to distil the knowledge of the upper model known as the teacher model to the smaller model recognised as the student model. This method was inspired by how people learn their daily tasks. The student model omits the teacher model's superfluous elements while continuously learning from it. The student model is more accurate and performs better than the teacher model, in the refinement of knowledge. The architecture is commonly used with models which have complex architecture. It is a computationally effective model.
2. **Generative Adversarial Network (GAN)** [31, 46] GANs are a type of unsupervised learning task of machine learning, which is trained using unlabeled data. Generator and Discriminator are the two neural network simulations that make up the GAN, as shown in Fig. 10 which automatically discover and learn the pattern from the input data. They compete with one another to study, record, and replicate the dataset's variances.

The Discriminator's job is to distinguish between authentic data and false data generated by the generator while the Generator in GAN generates fake data, or as we may say,

Table 3 Different models used in different studies

Research Study	Pre- processing	Segmentation	Task Classi- fication/ De- tection	Model
RS1	Y	Y	C	Class Activation Mapping with InceptionV3, ResNet50, Xception, DenseNet121 and EfficientNetB4
RS2	Y	Y	C	Multi-class SVM
RS3	Y	Y	C	U-Net + CNN
RS4	Y	N	C	Wavelet Transform based deep residual NN
RS5	Y	N	C	Knowledge Distillation
RS6	N	Y	C	Non-dominated Sorting Genetic Algorithm (NSGA II) + Pattern Net
RS7	Y	Y	C	CNN + Weakly supervised semantic segmen- tation algorithm
RS8	N	Y	D and C	ResNet 101, darkNet 19, DarkNet 53, ResNet, 50, GoogleNet
RS9	Y	N	C	GAN
RS10	Y	N	C	EfficientNet B6
RS11	Y	N	D	MobileNet + Squeeze-Mnet
RS12	Y	N	D	DenseNet 121 + MLP
RS13	Y	N	C	ResNet 50
RS14	Y	Y	D and C	Sparrow Search Algorithm + DenseNet201
RS15	Y	Y	D and C	AlexNet, VGGNet and ResNet
RS16	Y	Y	D and C	EfficentNets B4-B7
RS17	Y	N	C	DarkNet 53
RS18	Y	Y	C	ResNet 50
RS19	Y	N	C	DenseNet 201
RS20	Y	N	D and C	EfficientNetV2B0, B1 and B2
RS21	N	N	D	Feed Forward NN + Backpropagation Neural Network
RS22	Y	Y	C	Fractal NN based Galactic Swarm Optimization (FNN-GSO)
RS23	Y	Y	C	KNN
RS24	Y	Y	C	Swallow Swarm Optimization (SSO)

Table 3 (continued)

Research Study	Pre- processing	Segmentation	Task Classi- fication/ Detection	Model
RS25	N	N	D and C	MobileNetV2
RS26	Y	Y	D	Grey level co-occurrence matrix (GLCM) + Discrete wavelet transformation (DWT) + BPNN
RS27	N	Y	C	Moth Flame Optimization + Cubic SVM
RS28	Y	N	C	Fedrated Learning
RS29	Y	N	C	ResNet50
RS30	Y	N	C	MobileNetV3 large
RS31	Y	N	C	Inception ResNet v2
RS32	Y	Y	D	Aquila Whale Optimization + SqueezeNet
RS33	Y	N	C	Vision Transformer
RS34	Y	N	C	Random Forest
RS35	Y	Y	D and C	SqueezeNet + DNN
RS36	Y	Y	C	Bayesian optimized SegNet
RS37	Y	N	C	Enhanced Deep Bottleneck Transformer
RS38	Y	Y	C	ResNet50 + SVM
RS39	Y	N	C	VGG19 + CNN
RS40	Y	Y	D	DenseSFNet-45
RS41	Y	N	C	Image Visibility Filter + ResNet
RS42	Y	N	C	Deep Convolutional Neural Network
RS43	Y	Y	C	Mask Recurrent Convolutional Neural Network
RS44	Y	Y	C	SVM + RF
RS45	Y	Y	C	ResNet 50 + Feature pyramid Network + DenseNet201

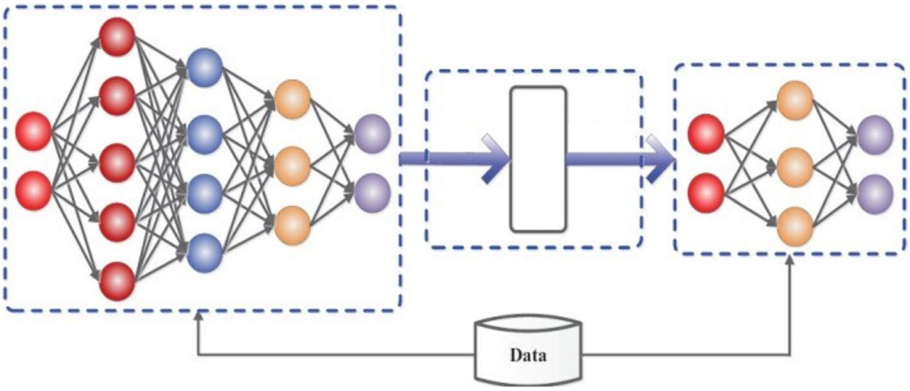


Fig. 9 Teacher-Student model of Knowledge distillation [49]

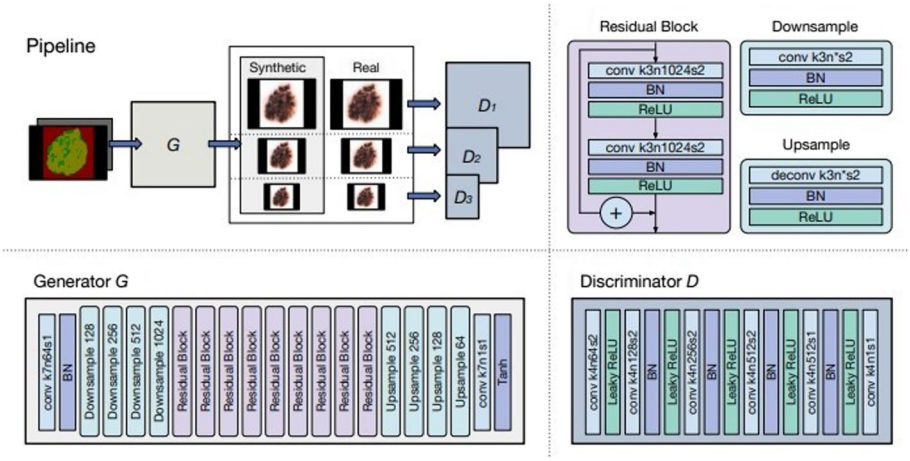


Fig. 10 GAN Architecture [16]

comparable data of the training set. The real data, which is used as a positive sample, and the generator's fictious data, which serves as a negative sample during training, are the two sources from which the discriminator data is derived.

They play an antagonistic game during training where the generator tries to mislead the discrim- inator by providing data that is similar to the data in the training set and the discriminator tries to avoid being tricked by being able to tell the difference between false and real data. To learn and train complicated data, like as videos, images, and audio, they both work simultaneously. Both the generator and the discriminator get better over time at generating realistic samples and recognising actual samples from phoney ones.

GANs have a considerably high potential for skin disease detection and classification by producing real samples that can be used to train ML models. To supplement sparse datasets and boost the precision of classification algorithms, GANs have been employed to create artificial skin lesion images. For example, the PH2 dataset only has 200 images which is very low for a deep-learning model. For this type of problem GANs can give a

solution by producing more synthetic and realistic samples so as to train the model on more datasets, which gives high accuracy.

One of the problem seen while training a dataset is overfitting. In the medical domain, controlling overfitting is essential to ensure that ML and DL models generalise well to new data and make accurate predictions. Some of the techniques that can address overfitting are:

1. Increase the data size
2. Data Augmentation
3. Feature Selection
4. Regularization
5. Drop Out
6. Early Stopping
7. Feature Engineering
8. Hyperparameters tunning

It's critical to select the best combination of these methods based on the particular medical dataset, the difficulty of the issue, and the resources at hand. In the medical sector, finding the ideal balance between model complexity and generalisation performance frequently requires iterative experimentation and extensive validation.

Research studies using ML or DL or hybrid are shown in Figure 11. Different ML and DL techniques which are used in recent 45 studies are explained in Table 3 which

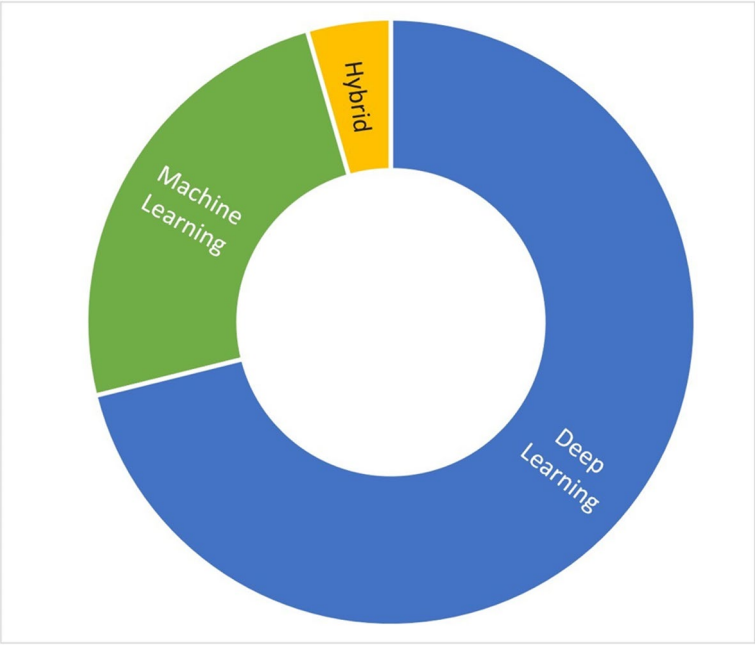


Fig. 11 Count of techniques

have explained the task of whether the study is detecting skin disease or classifying it into further diseases. Classification is abbreviated as 'C' and Detection as 'D'.

3.1.4 RQ4: Which data sets are being utilised to train the models?

Datasets are given as input to the model for training so that the model can train itself to classify skin diseases. Each study has used some datasets to train their models and to produce results. The datasets we found in the studies we are reviewing are mostly standard publicly available datasets and there are very few studies that have been used from other sources like from some hospital or Dermatology department, whether they might be publicly or not. Most of the researchers took real-time datasets by collecting images of healthy or diseased skin which were collected by some international collaboration.

In some of the datasets balancing the dataset is a common problem. According to the authors there are mainly three approaches by which we can handle the problem of dataset. First is Resampling which involves oversampling and undersampling. In oversampling duplicate the existing samples in the minority class and in undersampling we randomly remove the samples from the majority class. Second is Cost-sensitive learning in which we assign different weights to different classes during training. And last is data augmentation in which we generate new samples by transformations like rotation, scaling, cropping etc. to the existing data. The dataset used in each study is given in Table 4 and Figure 12. Some publicly and standard datasets are also mentioned which is not used by any of study like Dermis [33], DermNet[34], DermoFit[35] and EDRA [12] are also mentioned in Table 4 and Figure 12 for knowledge purpose. Another observation from the research studies is that most of the mentioned datasets are variable in size, and most datasets are large in size.

3.1.5 RQ5: What are the main performance measures used for validating skin-related diseases?

To assess the models' performance, different metrics are used to find if the model is performing better or not. These metrics are called performance metrics. Depending on different types of models, some of the popular and important result metrics are discussed below [44, 69]:

- Accuracy: The number of accurate predictions made by the model divided by the total number of predictions made is used to determine accuracy. The accuracy formula is:

$$\frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{1}$$

- Precision: The ratio of true positives to all of the correctly predicted positive outcomes is used to calculate the precision. For precision, use the formula:

$$\frac{T_p}{T_p + F_p} \tag{2}$$

- Sensitivity (Recall): The number of true positives divided by the total number of real positive cases is used to calculate recall. The recall equation is:

Table 4 List of popular dataset

Name	Year	No. of Images	Image Dimensions	Image Fomat	Used in	Research Study
Dermis[33]	NA	69	2067×1555	.jpeg	Melanoma (MEL), Nevusor or Atypical Nervus (NEV)	None
DermNet[34]	1998	23,000	540×722 to 4499×6748	.bmp	23 different types of diseases	None
PH2 [1]	2013	200	766×576	.bmp and.jpeg	MEL, NEV, Common Nevus (CN)	RS2, RS6, RS7, RS12, RS14, RS22, RS40, RS41
ISIC 2016[56]	2016	2094	6000×4000	.jpeg	NEV, MEL	RS11, RS22, RS25, RS31, RS32
ISIC 2017[57]	2017	2862	6000×4000	.jpeg and.png	NEV, MEL, Seborrheic Keratosis (SK)	RS1, RS4, RS11, RS12, RS15, RS22, RS25, RS26, RS31, RS32, RS37, RS41, RS42
ISIC 2018[58]	2018	10,015	781×704	.jpg	NEV, MEL, SK, Basal Cell Carcinoma (BCC)	RS1, RS2, RS11, RS16, RS20, RS22, RS25, RS31, RS32, RS42, RS44
ISIC 2019[59]	2019	25,331	841×759	.jpeg	NEV, MEL, SK, BCC, AK	RS2, RS5, RS10, RS11, RS16, RS20, RS22, RS25, RS31, RS32, RS39, RS42, RS44
(Paid) [35]	NA	1300	380×380	.jpeg	NEV, MEL, SK, BCC, AK	None
EDRA [12]	2000	1011	768×512	.jpeg	NEV, CN, MEL, SK, BCC	None
ISIC 2020[60]	2020	33,126	850×955 to 4499×6748	.jpg	NEV, CN, MEL, SK, BCC	RS10, RS11, RS20, RS22, RS31, RS32
Kaggle skin Cancer [38]	2018	1800	224×244	.jpg	Bengin and Malignant skin mole	RS9, RS14
HAM10000[70]	2018	10,015	400×600	.jpeg	AKIEC, MEL, BCC, BKL, DF, NV, VASC	RS2, RS3, RS4, RS9, RS12, RS18, RS19, RS20, RS22, RS27—RS30, RS33, RS37, RS38, RS43, RS45
ISBI 2016[86]	2016	900	768×500	.jpeg	NV, SK, MEL	RS36, RS45
ISBI 2017[64]	2017	2000	4500×6000	.jpeg	NV, SK, MEL	RS36, RS40, RS45
ISBI 2018[90]	2018	2594	4500×6000	.jpeg	AKIEC, BCC, BKL, DF, VASC, MEL	RS40
Others (Including Hospitals)	NA	Different	Different	Different	Various Diseases	RS8, RS13, RS17, RS21, RS23, RS28, RS34

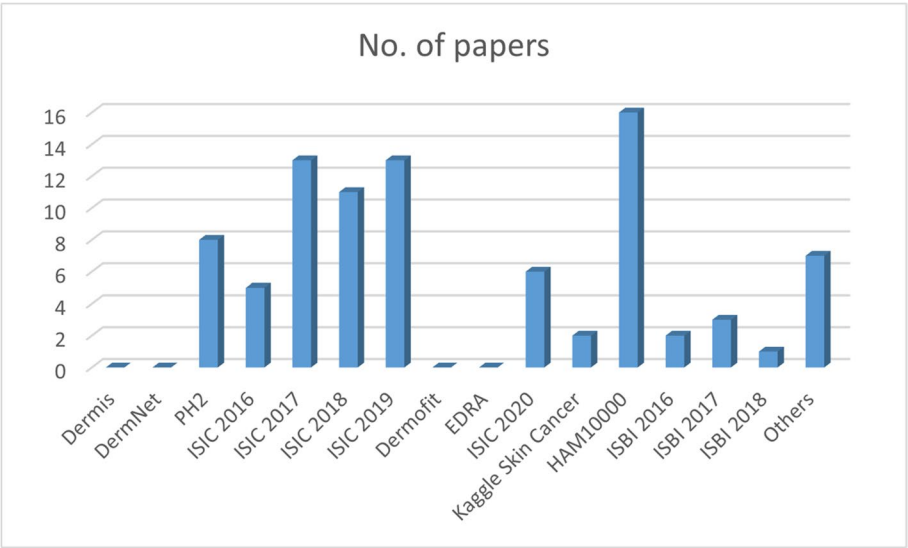


Fig. 12 Available datasets used for different studies

$$\frac{T_p}{T_p + F_n} \tag{3}$$

- Specificity: The number of true negatives divided by the total number of real negative cases is used to calculate specificity. The definition of specificity is:

$$\frac{T_n}{T_n + F_n} \tag{4}$$

- F1 Score: One interpretation of the F1 score is the harmonic mean of recall and precision, where the best value of it is 1 and the worst is 0. The formula:

$$2_* \frac{Precision * Recall}{Precision + Recall} \tag{5}$$

- AUC-ROC: AUC is Area under the curve and ROC is Receiver operating characteristics. AUC- ROC represents the probability of the model to be able to distinguish between two classes i.e. True positive and false positive.

  $(Tp : TruePositive, T_n : TrueNegative, F_p : FalsePositive, F_n : FalseNegative)$

The performance metrics used in different studies are shown in Table 5 and the count of studies using the specific performance metrics is given in Figure 13.

3.1.6 RQ6: How can we deal with noisy data from different sources?

Normally noise in an image is in various forms, it can be hairs, air bubbles, pen marks, oil marks blood vessels, scale marks, discoloration etc. which are shown in Figure 14. They

Table 5 Different performance metrics to measure the performance of models

Research Study	Accuracy	Specificity	Sensitivity/ Recall	Precision	F1 Score	ROC
RS1	93.5	-	-	-	-	-
RS2	98.98	-	-	-	-	-
RS3	97.96	-	-	-	-	-
RS4	96.91	97.68	-	96.43	95.79	-
RS5	84.6	-	-	-	-	-
RS6	92.24	100	100	-	-	-
RS7	93.02	-	-	-	-	-
RS8	90.75	-	-	-	-	-
RS9	98.89	-	-	-	-	-
RS10	-	-	-	-	-	99.48
RS11	99.36	-	-	-	-	-
RS12	98.33	-	-	-	-	-
RS13	87.42	-	-	-	-	-
RS14	94.16	-	-	-	-	-
RS15	95.56	96.23	97.98	-	-	-
RS16	94.13	-	-	-	-	-
RS17	97.14	-	96.49	96.5	-	-
RS18	86	-	84	86	86	-
RS19	97.16	98.88	92.76	97.45	-	-
RS20	99	-	99	99	99	-
RS21	86.04	-	-	-	-	-
RS22	98.4	97.2	98.01	97.42	97.4	-
RS23	98	-	-	-	-	-
RS24	98.5	99.1	94.5	-	-	-
RS25	93.9	-	-	-	-	-
RS26	93.9	-	-	-	-	-
RS27	95.9	-	-	-	-	-
RS28	92.76	-	-	-	-	-
RS29	88	-	74.42	71.28	72.1	-
RS30	90	-	81	86	86	99
RS31	89.55	-	91.67	-	-	-
RS32	92.5	91.7	92.1	-	-	-
RS33	94.1	-	-	94.1	94.1	-
RS34	90	-	-	93	94	-
RS35	99.9	99.94	99.49	98.6	99.03	-
RS36	95.1	95.84	90.89	-	-	-
RS37	95.84	-	-	96.1	-	-
RS38	99.87	-	98.87	98.77	97.83	-
RS39	96.91	-	92.18	92.18	92.19	92.18
RS40	97.7	-	-	-	-	-
RS41	96.8	98.2	97.9	95.8	-	-
RS42	93.25	90.64	93.25	93.97	93.47	-
RS43	86.5	-	85.57	87.01	86.28	-
RS44	89.31	87.81	88.56	-	-	-

Table 5 (continued)

Research Study	Accuracy	Specificity	Sensitivity/ Recall	Precision	F1 Score	ROC
RS45	93.6	-	96.25	96.44	96.34	-

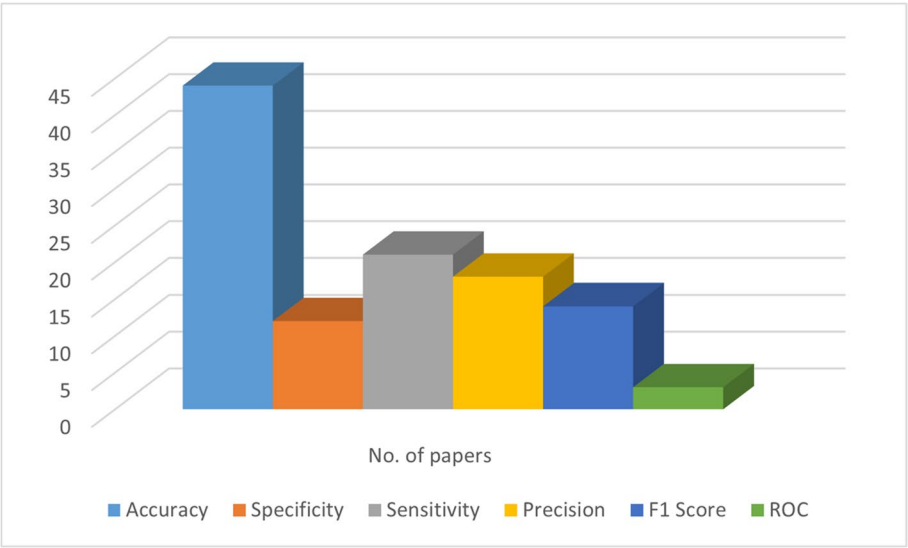


Fig. 13 Performance Metric

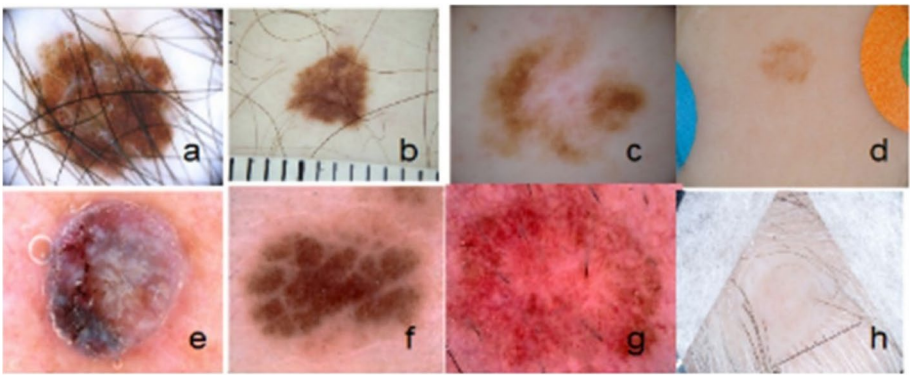


Fig. 14 Noises in skin lesion (**a**) hair artefact (**b**) ruler mark artefact (**c**) low contrast (**d**) color illumi- nation (**e**) bubbles (**f**) irregular boundaries (**g**) blood vessels (**h**) frame artefact [2]

can be deal with in preprocessing stage and later at the image segmentation stage. Some of these techniques include:

1. Image cropping
2. Data augmentation

3. Image resizing
4. Image normalization

Image segmentation is another type of pre-processing technique or we can say it is the advanced version of pre-processing which is discussed above. It seeks to classify related areas or portions of an image under their corresponding class names [45].

Techniques for segmenting data include edge detection, region expansion, and thresholding, among others [45]. Setting a threshold value for pixel intensity makes the pixels above that value part of the lesion and below that value separate from it. To create regions that match the lesion, adjacent pixels with identical intensity levels are iteratively grouped to generate groups. Edge identification entails locating the lesion's edges based on variations in pixel intensity. Authors of [103] use Water- shed segmentation to identify cell boundaries. However, this method has limitations in capturing cell boundaries, leading to poor performance with noisy images and dataset training. The Table 3 displays the preprocessing and segmentation methods used in each research paper. We can conclude from this whether or not preprocessing and segmentation techniques are beneficial. Will they contribute to the model's accuracy improvement or not? We have deduced from this study that preprocessing and segmentation enhance the model's accuracy while also assisting in the removal of noise. Figure 15 illustrates several different kinds of segmentation procedures. For the study which used any prepro- cessing and segmentation techniques we put Yes 'Y' in front of it otherwise No 'N' respectively shown in Figure 3.

Segmentation can improve the accuracy and efficiency of skin disease detection and classification when combined with other preprocessing methods and machine learning algorithms.

1. Thresholding [45]: Setting a pixel intensity threshold is known as thresholding. Above this threshold, pixels are regarded as being a part of the lesion, and below it, they are not. There are various thresholding techniques, such as Otsu's method, which determines the threshold value using the pixel intensity histogram.
2. Region growing [45]: In order to create regions that match to the lesion, adjacent pixels with similar intensity levels are iteratively grouped together. For segmenting lesions with clearly defined borders, this technique is quite helpful.

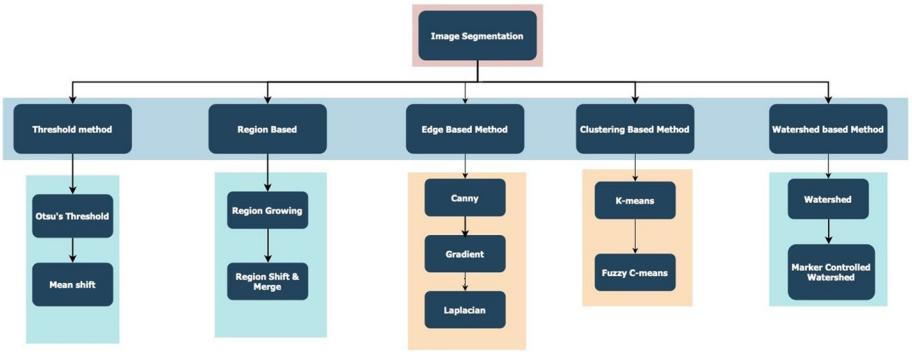


Fig. 15 Different types of Segmentation techniques [45]

3. Edge detection [45]: Edge detection includes locating the lesion's edges using variations in pixel intensity. There are several ways for detecting edges, such as Canny edge detection, which uses gradients in pixel intensity to do so.
4. Watershed segmentation [45]: Watershed segmentation is a method that can be used to divide lesions that are contacting or overlapping. It entails flooding the image from the highest spots to create basins around the lesions. The image is treated as a topographic map where the intensity values represent heights.
5. Deep Learning [99]: Skin lesion segmentation can also be done using deep learning methods like convolutional neural networks (CNNs). A big dataset of images can be used to train CNNs to recognise the characteristics of skin lesions.

The choice of segmentation technique depends on the specific dataset, the characteristics of skin lesion images and the aim of analysis. There is no one-fits-for-all answer, as the effectiveness of a technique widely depends on factors which are mentioned just before. In our research majority of research articles use segmentation techniques which boosts their results. In [102] authors find that segmentation increases the manual annotation cost and computation resource. In [73] uses CNN architecture for skin lesion segmentation, results are also improved. For PH2, it is from 98.21% to 98.98%, for ISIC 2019 it is 92.57% to 93.47%, for HAM10000 it is 95.22% to 98.57% and for ISIC 2018 it is 96.75% to 98.62%. In real-world situations, it can be helpful to test out several segmentation methods and gauge how well they function using your particular dataset. CNNs-based techniques, such as semantic or instance segmentation, have shown promise in automatically learning pertinent characteristics and carrying out segmentation in an end-to-end way. When you have a dataset that is sufficiently large and diverse, these approaches might perform better than conventional procedures.

3.1.7 RQ7: What can be the future research directions for this field?

ML and DL are showing great results for skin disease detection and classification but still, there are some challenges that need to be addressed to make it a broad clinical practice. One issue is that current models are frequently trained on biased and limited datasets. When the models are employed to categorise images from new patients, this can result in subpar results [63, 83]. To create models that are more resilient to these difficulties, more study is required. Another difficulty is that it's crucial to comprehend why a model generates a specific prediction. This is crucial in the medical industry since doctors need to have confidence in the model outcomes. To create techniques for making machine learning and deep learning models more comprehensible, more study is required.

Last but not least, a lot of DL and ML models demand a lot of data and computing power [36]. This can make it challenging to use these models in environments with scarce resources, like rural areas or underdeveloped nations. To create models that can be applied in these contexts, more research is required. Images are only one sort of data that may be utilised to increase the precision of skin disease identification and categorization. Other types of data, such as medical history and test findings, can also be employed. The utilisation of multimodal data in this area requires further study. Some other research can be used to make this more precise such as face detection, head pose classification [4] and mouth region extraction [5] which helps to recognize the skin-specific area, affected by a disease. Explore transfer learning approaches further so that models that have already been trained

on huge datasets can be customised for particular skin disease detection tasks. This strategy can help with the lack of data on rare diseases [15].

4 Conclusion

The research on skin lesion picture classification and disease recognition is analyzed in this paper. It is an extensive and little diverse but still concise survey of the techniques and procedures utilized in the handling, segmentation, and classification of skin lesion pictures. Both traditional machine learning methods and deep convolutional neural network models were investigated. The known and available datasets were discussed, and then these datasets were compared. The use of ML and DL to identify and classify skin diseases has shown great promise for automating and increasing the accuracy of skin disease diagnosis. When it came to deep learning, there were a number of ways to learn. Some of them used fully connected CNNs in combination with hybrid ML techniques, while others relied on ensemble approaches or transfer learning. The pre-trained model has demonstrated potentially encouraging results for high-accuracy melanoma detection and classification using the class of meta-heuristic optimisation approaches, which are inspired by the nature approach. We find that the non-parametric approach shows promising results in differentiating images, so it may also provide for the skin disease dataset as well.

Due to the small size of many of the datasets, there is a limited range of images for testing and preparation that are available for analysis, for instance, PH2 and Dermis have 200 and 69 images only. The proposed approaches perform well with small datasets, but they are susceptible to overfitting and reliably unpredictable when tested with a big image set. Small dataset preparation could be facilitated by data augmentation and images generated by GANs. Additionally, there is a way, when the datasets from healthcare or medical will be available publicly or can be used by researchers. A few specialists utilize non-public information bases and web pictures. Replicating the findings and conclusions is more challenging for the internet images due to bias in the dataset. Making big public image databases containing images that are as representative of the global population as feasible is a critical effort in this field of research in order to prevent racial bias. Picture bias in light of orientation and race artificial intelligence bias implies that the models and calculations neglect to give ideal outcomes for individuals of an under-addressed orientation or nationality. Most skin lesions on light-colored skin can be seen in the current databases. For instance, the images used for the ISIC dataset are often taken in the USA, Europe, and Australia. Likewise, CNN's attempt to separate the skin tone accomplishes a legitimate characterization of darker-looking people. This is only possible if the training dataset contains a sufficient number of images of individuals with dark skin. Another crucial element is the lesion's size. Assuming that the sore size is more modest than 6mm, melanoma cannot be distinguished without any problem. The expansion of clinical information, for example, race, age, orientation, and skin type, as contributions for classifiers might assist with expanding characterization precision. This supplemental information could be valuable for dermatologists' navigation. These viewpoints ought to be remembered for future work.

Lastly, if the dataset has a lot of images for each class, the authors contend that deep learning outperforms traditional machine learning. Based on how big the dataset is, this. Indeed, even with datasets containing not many pictures, profound learning can defeat this issue by utilizing various strategies for increase. Deep learning produces intelligent decisions independently and more accu- rately. Promising results have been seen for the high

accuracy exactness of melanoma identification using prepared profound learning models and customised strategies that relied on a profound learning approach.

Data bias, interpretability, and accessibility are just a few of the issues and limitations that need to be addressed. However, it is anticipated that these methods will become more widely used and incor- porated into clinical practice as the field develops, allowing for more effective and efficient treatment of skin diseases.

Funding No funding was received for this study.

Data availability The manuscript does not have any associated data that was generated.

Declarations

There are no studies by any of the authors in this article that used humans or animals as subjects.

Conflict of interest Author 1 declares that she has no conflict of interest.
Author 2 declares that she has no conflict of interest.

References

1. ADDI Project (2013) https://www.fc.up.pt/addi/ph2%20database.html. PH2 Dataset, Accessed January 30, 2023
2. Adegun A, Viriri S (2021) Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art. Artif Intell Rev 54:811–841
3. Devakishan Adla et al. (2022) Deep learning-based computer aided diagnosis model for skin cancer detection and classification. In: Distributed and Parallel Databases 40.4, pp. 717–736. https://doi.org/10.1007/s10619-021-07360-z
4. Sadia Afroze Md, Hossain Rajib, Hoque Mohammed Moshiul (2022) Deepfocus: A visual focus of attention detection framework using deep learning in multi-object scenarios. In J King Saud Univ-Comput Inf Sci 34(10):10109–10124
5. Afroze S et al. (2023) An empirical framework for detecting speaking modes using ensemble classifier. In: Multimed Tools Appl, pp. 1–34. https://doi.org/10.1007/s11042-023-15254-8
6. Alenezi Fayadh, Armghan Ammar, Polat Kemal (2023) Wavelet transform based deep residual neural network and ReLU based Extreme Learning Machine for skin lesion classification. In Exp Syst Appl 213:119064
7. Nawal Soliman ALKolifi ALEnezi (2019) A method of skin disease detection using image processing and machine learning. Proced Comput Sci 163:85–92
8. Almuayqil Saleh Naif, El-Ghany Sameh Abd, Elmogy Mohammed (2022) Computer-Aided Di- agnosis for Early Signs of Skin Diseases Using Multi Types Feature Fusion Based on a Hybrid Deep Learning Model. In: Electron 11.23:4009
9. Alwakid G, Gouda W, Humayun M, Sama NU (2022) Melanoma detection using deep learning-based classifications. Healthcare 10(2):2481
10. Anand Vatsala et al (2022) An automated deep learning models for classification of skin disease using Dermoscopy images: A comprehensive study. Multimed Tools Appl 8126:37379–37401
11. Anand Vatsala et al (2023) Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images. In: Exp Syst Appl 213:119230
12. Argenziano G, Soyer HP, De Giorgi V, Piccolo D, Carli P, Delfino M (2000) Interactive atlas of dermoscopy (book and CD-ROM)
13. Hossam Magdy Balaha and Asmaa El-Sayed Hassan (2023) Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm. Neural Comput Appl 351:815–853
14. Bandy AD et al. (2023) Intraclass clustering-based CNN approach for detection of malignant melanoma. In: Sensors 23.2, p. 926. https://doi.org/10.3390/s23020926
15. Catarina Barata M, Celebi Emre, Marques Jorge S (2018) A survey of feature extraction in der- moscopy image analysis of skin cancer. IEEE J Biomed Health Inf 233:1096–1109
16. Bissoto A, Perez F, Valle E, Avila S (2018) Skin lesion synthesis with generative adversarial networks. In: OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical

Image-Based Procedures, and Skin Image Analysis: First International Workshop, OR 2.0 2018, 5th International Workshop, CARE 2018, 7th International Workshop, CLIP 2018, Third International Workshop, ISIC 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16 and 20, 2018, Proceedings 5. Springer International Publishing, pp 294–302
17. Marta Bistron´ and Zbigniew Piotrowski (2022) Comparison of Machine Learning Algorithms Used for Skin Cancer Diagnosis. In: Appl Sci 12.19:9960
18. Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on Computational learning theory, pp 144–152
19. Breiman L (1996) Bagging predictors. Mach Learn 24:123–140
20. Breiman L (2001) Random forests. Mach Learn 45:5–32
21. Choudhary Priya, Singhai Jyoti, Yadav JS (2022) Skin lesion detection based on deep neural networks. In: Chemometrics Intell Lab Syst 230:104659
22. Mayo Clinic (2022) Actinic Keratosis. https://www.mayoclinic.org/diseases.conditions/actinic-keratosis/symptoms-causes/syc-20354969. Accessed May 15, 2023. December 17
23. Mayo Clinic (2023) Atopic dermatitis (eczema). https://www.mayoclinic.org/diseases-conditions/atopic-dermatitis-eczema/symptoms-causes/syc-20353273. Accessed May 15, 2023. May 09
24. Mayo Clinic (2021) Basal Cell Carcinoma. https://www.mayoclinic.org/diseases-conditions/.basal-cell-carcinoma/symptoms-causes/syc-20354187. Accessed May 15, 2023. October 01
25. Mayo Clinic (2022) Melanoma. https://www.mayoclinic.org/diseases-conditions/melanoma/symptoms-causes/syc-20374884.Melanoma, Accessed May 15, 2023. June 18
26. Mayo Clinic (2022) Psoriasis. https://www.mayoclinic.org/diseases-conditions/psoriasis/symptoms-causes/syc-20355840. Accessed May 15, 2023. October 08
27. Mayo Clinic (2021) Rosacea. https://www.mayoclinic.org/diseases.conditions/rosacea/symptoms-causes/syc-20353815. Accessed May 15, 2023. September 22
28. Mayo Clinic (2022) Seborrheic Keratosis. https://www.mayoclinic.org/diseases-conditions/seborrheic-keratosis/symptoms-causes/syc-20353878. Accessed May 15, 2023. January 18
29. Mayo Clinic (2022) Seborrheic keratosis. https://www.mayoclinic.org/diseases-conditions/seborrheic-keratosis/symptoms-causes/syc-20353878. Seborrheic Keratosis, Accessed May 15, 2023. January 18
30. Mayo Clinic (2021) Squamous Cell Carcinoma of the skin. https://www.mayoclinic.org/diseases.conditions/squamous-cell-carcinoma/symptoms-causes/syc-20352480. Accessed May 15, 2023. May 13
31. Creswell Antonia et al (2018) Generative adversarial networks: An overview. In IEEE Signal process Mag 351:53–65
32. Daghrir J, Tlig L, Bouchouicha M, Sayadi M (2020) Melanoma skin cancer detection using deep learning and classical machine learning techniques: a hybrid approach. In: 2020 5th international conference on advanced technologies for signal and image processing (ATSIP). IEEE, pp 1–5
33. Dermatology Information System (2023) https://www.dermis.net/dermisroot/en/home/index.htm/.DermIS.Dataset, Accessed January 30
34. DermNet-All about Skin (1998) https://dermnetnz.org/.DermNet.Dataset, Accessed January 30, 2023
35. DERMOFIT Project Datase (2023) https://homepages.inf.ed.ac.uk/rbf/DERMOFIT/datasets.htm.DermoFit Dataset, Accessed January 30
36. Dildar Mehwish et al (2021) Skin cancer detection: a review using deep learning techniques. In: Int J Environ Res Publ Health 1810:5479
37. Diwan Tausif et al (2023) Model hybridization & learning rate annealing for skin cancer detection. In: Multimed Tools Appl 822:2369–2392
38. Fanconi C (2018) Skin Cancer: Malignant Vs Benign. https:/wwwkaggle.com/ datasets/fanconic/skin-cancer-malignant-vs-benign.Skin Cancer, Accessed January 30, 2023
39. Farag A et al. (2016) A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling. In: IEEE Trans Image Process 26.1, pp. 386–399. https://doi.org/10.1109/TIP.2016.2624198
40. Gajera Himanshu K, Nayak Deepak Ranjan, Zaveri Mukesh A (2023) A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features. In: Biomed Signal Process Control 79:104186
41. Ganguly Biswarup, Dey Debangshu, Munshi Sugata (2021) Image visibility filter-based inter- pretable deep learning framework for skin lesion diagnosis. IEEE Trans Industr Inf 18.8:5138–5147
42. Ghahfarrokhi SS et al. (2023) Malignant melanoma diagnosis applying a machine learning method based on the combination of nonlinear and texture features. In: Biomedical Signal Processing and Control 80, p. 104300. https://doi.org/10.1016/j.bspc.2022.104300
43. Goceri E (2021) Deep learning-based classification of facial dermatological disorders. In: Comput Biol Med 128, p. 104118. https://doi.org/10.1016/j.compbiomed.2020.104118

44. Goceri E (2021) Diagnosis of skin diseases in the era of deep learning and mobile technology. In: Computers in Biology and Medicine 134, p. 104458. https://doi.org/10.1016/j.compbiomed.2021.104458
45. Gonzalez RC, Woods RE (2008) Digital Image Processing. Prentice Hall. ISBN: 9780131687288. URL: https://books.google.co.in/books?id=8uGOnjRGEzoC. Accessed 12 May 2023
46. Goodfellow I et al. (2020) Generative adversarial networks. In: Communications of the ACM 63.11, pp. 139–144. https://doi.org/10.1145/3422622
47. Hatem MQ (2022) Skin lesion classification system using a K-nearest neighbor algorithm. In: Visual Computing for Industry, Biomedicine, and Art 5.1, pp. 1–10. https://doi.org/10.1186/s42492-022-00103-6
48. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
49. Hinton G, Vinyals O, Dean J (2015) Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531
50. Hong Y et al. (2023) Weakly supervised semantic segmentation for skin cancer via CNN super- pixel region response. In: Multimedia Tools and Applications 82.5, pp. 6829–6847. https://doi.org/10.1007/s11042-022-13606-4
51. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
52. Hum YC et al. (2022) The development of skin lesion detection application in smart hand- held devices using deep neural networks. In: Multimedia Tools and Applications 81.29, pp. 41579–41610. https://doi.org/10.1007/s11042-021-11013-9
53. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K (2016) SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360
54. Syed Inthiyaz et al. "Skin disease detection using deep learning". In: *Advances in Engineering Software* 175 (2023), p. 103361.
55. Iqbal Imran et al (2021) Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images. In: Computer Med Imaging Graph 88:101843
56. ISIC Challenge (2016) https://challenge.isic-archive.com/data/#2016. ISIC 2016 Dataset, Accessed January 30, 2023
57. ISIC Challenge (2017) https://challenge.isic-archive.com/data/#2017. ISIC 2017 Dataset, Accessed January 30, 2023
58. ISIC Challenge (2018) https://challenge.isic-archive.com/data/#2018. ISIC 2018 Dataset, Accessed January 30, 2023
59. ISIC Challenge (2019) https://challenge.isic-archive.com/data/#2019. ISIC 2019 Dataset, Accessed January 30, 2023
60. ISIC Challenge (2020) https://challenge.isic-archive.com/data/#2020. ISIC 2020 Dataset, Accessed January 30, 2023
61. Bethanney Janney J, Emalda Roslin S (2020) Classification of melanoma from Dermoscopic data using machine learning techniques. In: Multimed Tools Appl 79:3713–3728
62. Karuppiah SP et al (2022) An Efficient Galactic Swarm Optimization Based Fractal Neural Network Model with DWT for Malignant Melanoma Prediction. In: Neural Process Lett 54.6:5043–5062
63. Kassem Mohamed A et al (2021) Machine learning and deep learning methods for skin lesion classification and diagnosis: a systematic review. In: Diagnostics 11.8:1390
64. Khan MA et al. (2021) "Attributes based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework". In: Pattern Recognition Letters 143, pp. 58–66 https://doi.org/10.1016/j.patrec.2020.12.015
65. Khan MA et al. (2021) Pixels to classes: intelligent learning framework for multiclass skin lesion localization and classification. In: Comput Electric Eng 90, p. 106956. https://doi.org/10.1016/j.compeleceng.2020.106956
66. Krizhevsky Alex, Sutskever Ilya, Hinton Geoffrey E (2017) Imagenet classification with deep convolutional neural networks. In: Commun ACM 60.6:84–90
67. Anup Kumar K, Vanmathi C (2022) Optimization driven model and segmentation network for skin cancer detection. In: Comput Electric Eng 103:108359
68. LeCun Y et al. (1998) Gradient-based learning applied to document recognition. In: Proceedings of the IEEE 86.11, pp. 2278–2324. https://doi.org/10.1109/5.726791
69. Li Ling-Fang et al (2020) Deep learning in skin disease image recognition: A review. In: IEEE Access 8:208264–208280

70. Mader KS (2018) Skin Cancer MNIST: HAM10000. https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000. Skin Cancer, Accessed January 30, 2023
71. Malibari AA et al. (2022) Optimal deep neural network-driven computer aided diagnosis model for skin cancer. In: Comput Electric Eng 103, p. 108318. https://doi.org/10.1016/j.compeleceng.2022.108318
72. Manimurugan S (2023) Hybrid high performance intelligent computing approach of CACNN and RNN for skin cancer image grading. In: Soft Computing 27.1, pp. 579–589. https://doi.org/10.1007/s00500-022-06989-x
73. Maqsood S, Damaˇseviˇcius R (2023) Multiclass skin lesion localization and classifi- cation using deep learnin based features fusion and selection framework for smart healthcare. Neural Netw 160:238–258
74. Meena K et al (2022) A novel method for prediction of skin disease through supervised classification techniques. In: Soft Comput 26.19:10527–10533
75. Sufiyan Bashir Mukadam and Hemprasad Yashwant Patil (2023) Skin Cancer Classification Framework Using Enhanced Super Resolution Generative Adversarial Network and Custom Convo- lutional Neural Network. In: Appl Sci 13.2:1210
76. Murugan A et al (2021) Diagnosis of skin cancer using machine learning techniques. In: Microprocessors Microsyst 81:103727
77. Naeem Ahmad et al (2022) SCDNet: A Deep Learning-Based Framework for the Multiclassification of Skin Cancer Using Dermoscopy Images. In: Sensors 22.15:5652
78. Nakai Katsuhiro, Chen Yen-Wei, Han Xian-Hua (2022) Enhanced deep bottleneck transformer model for skin lesion classification. In: Biomed Signal Process Control 78:103997
79. Nguyen Viet Dung, Bui Ngoc Dung, Do Hoang Khoi (2022) Skin Lesion Classification on Im- balanced Data Using Deep Learning with Soft Attention. In: Sensors 2219:7530
80. Oakley A (2020) Dermatofibroma. https://dermnetnz.org/topics/dermatofibroma. Accessed May 15, 2023. September
81. Oakley A (2016) Melanocytic naevus. https://dermnetnz.org/topics/melanocytic.naevus. Accessed May 15, 2023. January
82. Oakley A (2016) Vascular proliferations and abnormalities of blood vessels. https://dermnetnz.org/topics/vascular-proliferations-and-abnormalities-of-blood-vessels. Accessed May 15, 2023. February
83. Pacheco AGC, Krohling RA (2019) Recent advances in deep learning applied to skin cancer detection. arXiv preprint arXiv:1912.03280
84. Park Seungman et al (2023) FACES: A Deep-Learning-Based Parametric Model to Improve Rosacea Diagnoses. In: Appl Sci 13.2:970
85. Ravi Vinayakumar (2022) Attention Cost-Sensitive Deep Learning-Based Approach for Skin Cancer Detection and Classification. In: Cancers 14.23:5872
86. Nurullah S¸ahin, Nuh Alpaslan, and Davut Hanbay (2022) Robust optimization of SegNet hyperparameters for skin lesion segmentation. In: Multimed Tools Appl 81.25:36031–36051
87. Salma Wessam, Eltrass Ahmed S (2022) Automated deep learning approach for classification of malignant melanoma and benign skin lesions. In: Multimed Tools Appl 81.22:32643–32660
88. Schapire Robert E (1990) The strength of weak learnability. In: Mach Learn 5:197–227
89. Sevli Onur (2021) A deep convolutional neural network-based pigmented skin lesion classification application and experts evaluation. In: Neural Comput Appl 33.18:12039–12050
90. Shan Pufang et al (2022) Automatic skin lesion classification using a new densely connected convolutional network with an SF module. In: Med Biol Eng Comput 60.8:2173–2188
91. Sharafudeen Misaj (2023) Detecting skin lesions fusing handcrafted features in image network ensembles. In: Multimed Tools Appl 82.2:3155–3175
92. Shen S et al. (2022) A low-cost high-performance data augmentation for deep learning-based skin lesion classification. In: BME Frontiers 2022. https://doi.org/10.34133/2022/9765307
93. Shinde RK et al. (2022) Squeeze-MNet: Precise Skin Cancer Detection Model for Low Com- puting IoT Devices Using Transfer Learning In: Cancers 151, p. 12. https://doi.org/10.3390/cancers15010012
94. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
95. Sun Nana et al (2022) Novel neural network model for predicting susceptibility of facial post- inflammatory hyperpigmentation. In: Med Eng Phys 110:103884
96. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826

97. Tan M, Le Q (2019) Efficientnet: rethinking model scaling for convolutional neural networks. In: International conference on machine learning. PMLR, pp 6105–6114
98. Thapar P et al. (2022) A novel hybrid deep learning approach for skin lesion segmentation and classification. In: J Healthcare Eng 2022. https://doi.org/10.1155/2022/1709842
99. Topiwala A, Al-Zogbi L, Fleiter T, Krieger A (2019) Adaptation and evaluation of deep learning techniques for skin segmentation on novel abdominal dataset. In: 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE). IEEE, pp 752–759
100. Muhammad Usama M, Naeem Asif, Mirza Farhaan (2022) Multi-Class Skin Lesions Classification Using Deep Features. In: Sensors 22.21:8311
101. Verma Anurag Kumar, Pal Saurabh, Kumar Surjeet (2020) Prediction of skin disease using ensemble data mining techniques and feature selection method—a comparative study. In: Appl Biochem Biotechnol 190:341–359
102. Wang L, Zhang L, Shu X, Yi Z (2023) Intra-class consistency and inter-class discrimination feature learning for automatic skin lesion classification. Med Image Anal 85:102746
103. Wang Y, Wang W, Liu D, Hou W, Zhou T, Ji Z (2023) GeneSegNet: a deep learning framework for cell segmentation by integrating gene expression and imaging. Genome Biol 24(1):235
104. Wang W, Han C, Zhou T, Liu D (2022) Visual recognition with deep nearest centroids. arXiv preprint arXiv:2209.07383
105. Wang Yongwei et al (2023) Ssd-kd: A self-supervised diverse knowledge distillation method for lightweight skin lesion classification using dermoscopic images. In: Med Image Anal 84:102693
106. Wicaksana Jeffry et al (2022) Customized Federated Learning for Multi-Source Decentralized Medical Image Classification. In: IEEE J Biomed Health Inf 26.11:5596–5607
107. Wolpert David H (1992) Stacked generalization. In: Neural Netw 5.2:241–259
108. Wu ZHE et al (2019) Studies on different CNN algorithms for face skin disease classification based on clinical images. IEEE Access 7:66505–66511
109. Xin Chao et al (2022) An improved transformer network for skin cancer classification. In: Comput Biol Med 149:105939
110. Yang L, Wang Q, Wang J, Quan X, Feng F, Chen Y, Khabsa M, Wang S, Xu Z, Liu D (2023) MixPAVE: mix-prompt tuning for few-shot product attribute value extraction. In: Findings of the Association for Computational Linguistics: ACL 2023, pp 9978–9991

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

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

Advanced Intelligent Hybrid Approach for State of Charge Estimation of Li-Ion Batteries in Electric Vehicles under Various Drive Cycles

Deepak Kumar[1,2]
Center for Electrical Vehicles and Related Technology, Department of Electrical Engineering, Delhi Technological University, Delhi, India
deepakvns163@gmail.com

M. Rizwan[1]
Center for Electrical Vehicles and Related Technology, Department of Electrical Engineering, Delhi Technological University, Delhi, India
rizwan@dce.ac.in

Amrish K. Panwar[2]
Lithium-Ion Battery Technology Laboratory, Department of Applied Physics, Delhi Technological University, Delhi, India
amrish.phy@dtu.ac.in

2023 IEEE International Conference on Power Electronics, Smart Grid, and Renewable Energy (PESGRE) | 979-8-3503-1057-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/PESGRE58662.2023.10404810

***Abstract*— Lithium-ion batteries are currently used in various fields for environmental and resource concerns. An accurate estimation of the state of charge (SOC) facilitates the battery's functioning effectively. Battery management systems (BMS), more frequently used in electric vehicles (EV) and smart grid technologies, need to estimate their SOC to run safely and efficiently. However, the issue of SOC estimation at ambient temperatures with different drive cycles has not received much attention from existing approaches. In this research, an advanced intelligent hybrid approach consisting of stacked long short-term memory (SBLSTM) networks and convolutional neural networks (CNN) was presented. The features are automatically extracted by using a two-dimensional convolutional neural network model which has been developed. In addition, an estimator called an SBLSTM captures the backward and forward temporal dependency of battery sequential states. The combined CNN-SBLSTM network is studied. SOC estimation is conducted under different drive cycles to check the network acceptability. The root mean square error and maximum SOC error are both estimated to be less than 0.1% and 0.2%, respectively. Furthermore, the computational time is achieved at 23.083 and 13.531 seconds. The results of the experiments demonstrate that the approach is more accurate, stable, and computationally efficient.**

Keywords- Electric Vehicle, Lithium-ion battery, Drive Cycle, State of charge.

I. INTRODUCTION

Air pollution has increased in recent decades due to the development of industry and the usage of vehicles that run on fossil fuels. Fossil fuel burning in vehicles releases pollutants including carbon monoxide, nitrogen oxides, and hydrocarbons into the atmosphere, which have been a major source of concern across the world [1]. In this reference, the automobile sector, increasing the production of electric vehicles, therefore, it has prompted academics to concentrate on battery management systems [2]. Due to their high energy density, high power density, and low self-discharge rate, lithium-ion batteries (LIBs) have drawn tremendous interest in the electronic and automotive sectors than other kinds of batteries [3].

The SOC, which is a key battery characteristic, accurately measures the battery's remaining electrical energy. An accurate SOC estimate is important for the BMS to determine the remaining useful life (RUL), and over-discharging and overcharging may be prevented based on an accurate SOC estimation[4]. Thus, a growing lot of research and the incorporation of several battery-based SOC investigations were spurred by the role and significance of the SOC in BMS [5]. The estimation of SOC is a complex issue as with continuous deceleration and acceleration of electric vehicles, the batteries experience many different dynamics. The state of charge cannot be determined directly; consequently, it is important to carefully estimate the SOC [6]. The four categories of SOC estimation approaches are open-circuit voltage, coulomb counting, model-based, and data-driven. The first approach is recognized for having shortcomings and outcomes during offline measurements that are inappropriate for true SOC estimation. The Coulomb counting method works effectively when battery capacity and initial SOC are known, as well as when the current is correctly computed. There are three serious concerns about applying the Coulomb counting approach for computing actual SOC; initial SOC must be known, battery capacity must be updated because of the battery aging, and inaccuracies during measurements were unavoidable.

The equivalent circuit, electrochemical, and electrochemical impedance models are the three most commonly used models for the model-based estimation method [7]. The internal parameters of the LIBs were calculated online using the battery's electrical and electrochemical equations in this method, which adopted techniques like the family Kalman filters (KFs), Luenberger observer, and H-observer. The accuracy of a model-based approach is dependent on the physical model and its internal characteristics, including internal resistance and capacitance, which is one of its main limitations. Such approaches were generally computationally inefficient and typically involve additional parameters or several models to execute SOC estimation during different circumstances [8].

Since data-driven estimation methods do not rely on internal battery parameters and instead use battery variables like temperatures, voltages, and current as inputs to train their models to estimate SOC [9]. They show significant advantages over traditional model-driven estimation methods [10]. Artificial neural networks, support vector machines, extreme learning machines, and others are some of the most popular artificial intelligence techniques. In contrast, the process of training takes longer when there is a lot of data available. Recurrent neural network (RNN) based techniques have outperformed the rest of the available techniques when applied to time series issues. In these networks, historical data is utilized to estimate future data. This would be attributed to the existence of memory cells in the architecture of such networks. Numerous studies have been conducted implementing such approaches in this study area, even though the previous SOC calculation in the SOC estimation study considerably influences the current SOC battery. The LSTM network, for instance, is excellent at collecting long-term dependencies from time series data and eliminates the gradient vanishing difficulty that arises in RNN from occurring in the

979-8-3503-1057-3/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:12:06 UTC from IEEE Xplore. Restrictions apply.

traditional back-propagation training procedure. Chemali et al. [11] suggested an LSTM-RNN network for SOC estimation while it trained the proposed network in 4 hours. Yang et al. [12] proposed a gated recurrent unit recurrent neural network (GRU-RNN), while GRU can handle the phenomenon known as gradient vanishing, for estimating LiFePO4 battery SOC. Furthermore, findings showed that it was capable of reliably estimating the SOC at a variety of temperatures but it takes 2.3 hours of training time. Zhang et al. [13] used bidirectional long short-term memory neural networks for estimating the SOC of LIBs in this investigation. Contrary to conventional unidirectional LSTM, the bidirectional LSTM can capture the forward and backward temporal dependencies of battery sequential data.

Convolutional neural networks (CNNs) are another important class of deep learning techniques for handling time series and visual data. To estimate SOC, the authors use a hybrid CNN-LSTM network, in which the CNN extracts spatial characteristics from current inputs and the LSTM network extracts temporal data from previous inputs [13]. The combination of these two methods provides a superior remedy for time series prediction issues. As a consequence, whereas CNN uses convolutional filters to extract relationships between input data, the Bi-LSTM describes long-term dependence and is great at handling time series data [14].

This paper proposed a combined CNN-SBLSTM model to represent complicated battery dynamics. By replacing the original signals with shorter feature vector sequences, the 2D-CNN is used to speed up the stacked BLSTM training process. It is also used to extract sophisticated spatial features from the original data. To simulate correlations between the past inputs and the present SOC, the stacked BLSTM is often used. To capture the temporal and spatial features of battery data, the proposed network combines CNN and stacked BLSTM networks.

These are the goals this article is trying to achieve.

- To demonstrate the effectiveness of CNN-SBLSTM to generalize by estimating the SOC on different driving cycles.

The remainder of this paper is organized as follows. The suggested CNN-SBLSTM architecture for SOC estimation is highlighted in Section II. The data set and drive cycles used in this investigation are described in Section III. discusses the achieved results using the CNN-SBLSTM and experimental setup in Section IV. The article is concluded in Section V.

II. State of Charge Estimation using Combined CNN-SBLSTM Models

The CNN-SBLSTM model foundational concepts and approach architecture are presented in this section.

A. Convolutional Neural Network

There are three different types of layers used for building CNN, such as convolutional, pooling, and fully connected layers [15]. The convolutional layer consists of a number of feature detector filters that are convolved using the input signal. A user-configurable stride value governs the filter's shifting distance relative to the signal. The generated feature maps are susceptible to down-sampling for dimension reduction to lessen overfitting via the pooling layer and processing power [16]. The last layer in conventional CNNs is the fully connected layer in most of CNNs, but in our model feature maps, as shown in Fig. 1, and directly goes to the stacked Bi-LSTM. In our investigation, three maximum pooling layers and three 2D-convolutional layers are used alternately. In each layer, batch normalization and the ReLU activation feature were utilized, which was proposed for nonlinearity and its strength in deep neural networks.

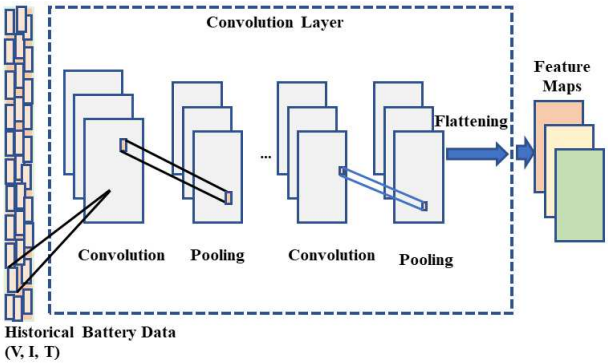


Fig.1: 2D-CNN model structure for feature extraction.

Recurrent neural networks (RNNs) are a type of neural network that can maintain its state across different inputs. It can accurately interpret a sequence or temporal series of data points without the need for the complete sequence to be given at once. Recurrent Neural Networks (RNNs) are highly effective in handling time sequences, such as battery signals including voltage, temperature, and current, as they have the ability to maintain states. When RNNs are trained by back-propagation, they experience bursting and vanishing gradients as well as information transformation [17]. The concept of memory cells (units) with adjustable gates was first introduced by LSTM. When back-propagating during training, it aids in sustaining gradients while preserving long-term temporal connections between inputs. As shown in Fig. 2, LSTM transition equations and a single LSTM cell's structure are as follows:

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \quad (1)$$

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \quad (2)$$

$$f_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) \quad (3)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh W_c x_t + U_c h_{t-1} + b_c \quad (4)$$

$$f_t = o_t \odot \sigma\left(c_t\right) \quad (5)$$

The sigmoid function σ, W signifies weight, while bias is b. The input gate i_t, while o_t is denoted as the output gate. The forget gate, memory cell, and hidden state are suggested by f_t, c_t, and, h_t respectively [18]. Schuster and Paliwal [19] developed bidirectional RNN a kind of RNN that used past and future input data sequences to train the network. The input data is processed using two linked layers and each layer performs operations in the reversed time step direction. The outcomes may be merged using a variety of merging strategies. To follow this principle, bidirectional LSTM involves two layers. The first layer performs tasks in the same direction as the data sequence, while the second layer carries out operations in the opposite direction.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:12:06 UTC from IEEE Xplore. Restrictions apply.

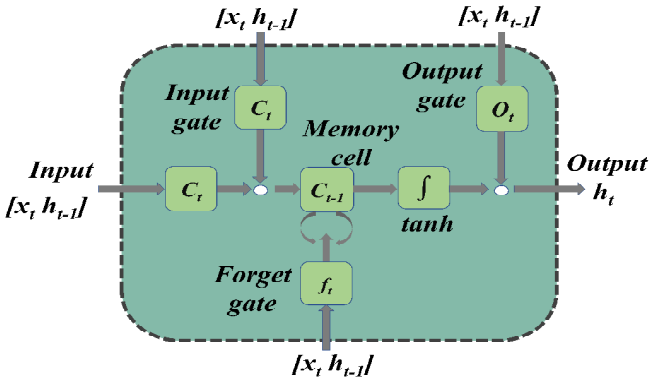


Fig. 2: LSTM model structure.

The BLSTM shows more accurate results than unidirectional LSTM in some applications, including battery management systems [20]. The unfolded BLSTM network topology is depicted in Fig. 3. There are two hidden layers in the BLSTM: the forward LSTM layer and the backward LSTM layer. Furthermore, both of the concealed layers are linked to the identical output layer. The forward LSTM layer output $\vec{h}_t$ is calculated by BLSTM independently, and then $\overleftarrow{h}_t$ every time step t is similarly concatenated through BLSTM. The LSTM update is represented in Eqs. (6-8).

$$\vec{h}_t = LSTM(x_t, \vec{h}_{t-1}) \quad (6)$$

$$\overleftarrow{h}_t = LSTM(x_t, \overleftarrow{h_{t+1}}) \quad (7)$$

$$y_t = W_{\vec{h}_y}\vec{h}_t + W_{\overleftarrow{h}_y}\overleftarrow{h}_t + b_y \quad (8)$$

where $W_{\vec{h}_y}$ indicates the weight from the forward LSTM layer to the output layer, $W_{\overleftarrow{h}_y}$ represents the weight from the backward LSTM layer to the output layer and the bias at the output layer is stated via by.

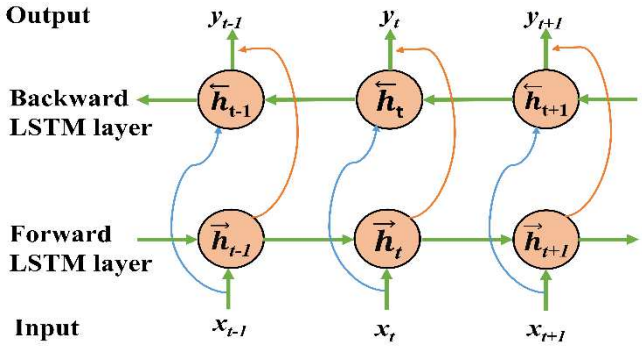


Fig. 3: Bidirectional LSTM model structure.

C. *Stacked Bidirectional LSTM Network*

In this paper, aim to use a stacked bidirectional LSTM network, illustrated in Fig. 4, to incorporate past and future sequential data in different directions and time intervals. The key role of our research is the proposal of a stacked BLSTM model, which allows for deeper and more accurate modeling by stacking the hidden layers. This model is applied as a deep learning technique for classification. The efficiency of neural networks in a wide range of challenging prediction problems is commonly known to describe their depth [21]. The input vector is ψ_t = [V (t), I(t), T(t)] where V (t), I(t), and T(t) denote the voltage, current, and temperature of the battery at time interval t in this proposed approach. At each time interval, the data is converted into an image and then fed into a 2D-CNN model to extract features. Once the convolutional layers have extracted the features, SBLSTM is used to train the feature maps by taking them as input. Lastly, the proposed technique estimated SOC at time t. The model includes two hidden layers, each with 25 nodes, using Bi-LSTM. Additionally, a fully connected (FC) layer and an output layer have been incorporated to enable the linear translation of the Bi-LSTM layers' hidden state into a single SOC output value. The output layer employs the rectified linear unit (ReLU) activation function. Furthermore, two stacked layers are used. As the model loss during each forward propagation, the mean square error (MSE) is determined by comparing the estimated SOC value with the true SOC value [22].

$$Loss = \frac{1}{m}\sum_{t=1}^{m} | (\hat{SOC}(t) - SOC(t)) |^2 \quad (9)$$

Here, the measurement capacity SOC (t), estimated capacity($\hat{SOC}(t)$, and m indicates the sample number.

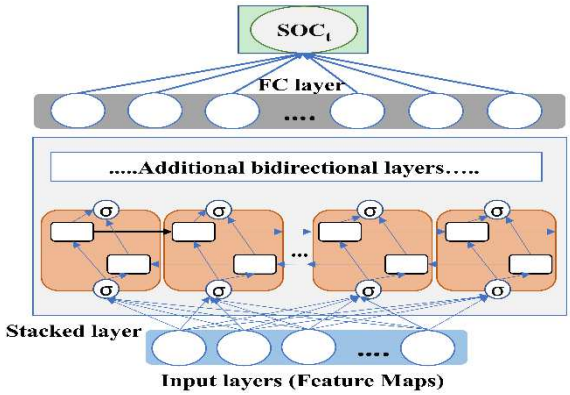


Fig. 4: Stacked Bidirectional LSTM model structure.

III. DATASET DESCRIPTION

A. *18650 Battery Datasets and Drive Cycles*

The chroma 17020 regenerative battery test system used as an experimental platform in this investigation is depicted in Fig. 4. In addition, all the battery tests are performed with the above battery tester. The host computer is used to record battery information, such as voltage, current, temperature, and so on. In these experiments 18650 lithium-ion batteries with 2.6 Ah rated capacity are studied. The test battery parameters are tabulated in Table 1. Based on the current profile of driving cycles, four different drive cycles are used. The Urban Dynamometer Driving Schedule (UDDS), Los Angeles 92 (LA92), Supplemental Federal Test Procedure Driving Schedule (US06), and Highway Fuel Economy Driving Schedule (HWFET) are published by the United States Environmental Protection Agency (EPA) publishes as shown in Fig. 5. The UDDS known for the light vehicle on city driving conditions and the LA92 simulates energy-efficient and low emission driving situations. The US06 simulates complex driving situations on the highway and HWFET signifies 77 km/h is the average speed at which driving conditions are simulated. These driving cycle profiles cover the battery load behaviors of electric vehicles in various environments. It should be clarified that all charge and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:12:06 UTC from IEEE Xplore. Restrictions apply.

discharge measurements are performed at 25 degrees Celsius at ambient temperatures.

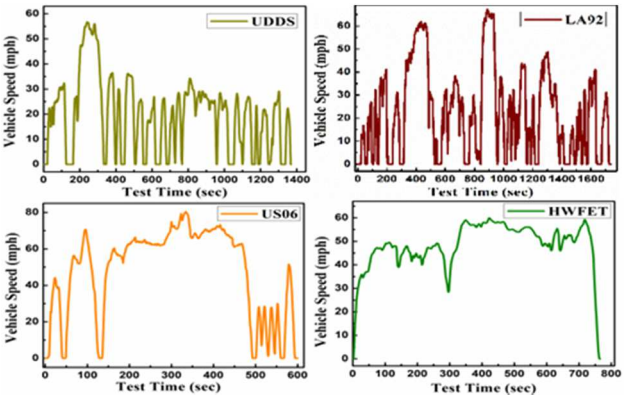


Fig. 5. The four drive cycles.

Table. I. Battery parameter specification.

Parameters	Value
Nominal voltage	3.6 V
Nominal capacity	2.6 Ah
Max/min voltage	4.2 / 2.5 V
Mass/energy storage	48 g / 9.9 Wh
Charging temperature (Minimum)	10 °C

IV. SOC ESTIMATION RESULTS AND ANALYSIS

A. Evaluation metrics

Three metrics are used to evaluate the estimation accuracy of the proposed CNN-SBLSTM network such as maximum absolute error (MaxError), mean absolute error (MAE), and root mean square errors (RMSE) are also used to evaluate comparisons with past research Eq. (10-13). The accuracy of the estimation is validated by MAE, while the robustness of the estimation by RMSE and the reliability of the estimation demonstrates MaxError.

$$MAE = \frac{1}{m}\sum_{t=1}^{m} | (\hat{SOC}(t) - SOC(t)) | \quad (10)$$

$$MSE = \frac{1}{m}\sum_{t=1}^{m}(\hat{SOC}(t) - SOC(t))^2 \quad (11)$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{t=1}^{m}(SOC_k - SOC'_k)^2} \quad (12)$$

$$MaxE = max_{t=1,\ 2,\ 3....m} | (\hat{SOC}(t) - SOC(t)) | \quad (13)$$

B. Experimental Details

The experimental work has been performed in our laboratory Centre of Excellence for Electric Vehicles and Related Technologies, Delhi Technological University, Delhi. The charge and discharge of the battery cells are controlled by Chroma regenerative battery pack test system model 17020 and its software as shown in Fig. 6. It has capable of charging and discharging the test battery in a range of preset current or voltage settings. The host computer containing Chroma 17020 software configures several settings, including sampling timing, cut-off necessities, and battery charging and discharging specifications. The battery negative and positive terminals are connected to the acquisition and output terminal of the Chroma 17020 test system. After that, the current



Fig. 6. Schematic diagram of the test bench and data logging system.

profile of the drive cycle, UDDS, LA92, US06, and HWFET is set on the host computer to perform the test and analyze the results. To record the battery surface temperatures by the inbuild chroma have been sensors. To build the program for the model training and simulation Python 6.4.12 were used. To train the CNN-SBLSTM model, the data of the tested battery was applied. The best hyperparameters of the network for the training were selected precisely.

C. SOC Estimation Results

To verify the accuracy of the proposed algorithm CNN-SBLSTM for the SOC estimation, this paper used four different drive cycles as shown in Fig. 5. The estimation performances of CNN-SBLSTM on different four drive cycles at 25˚C are shown in Fig. 7-10. In all four driving cycles, it performs well. The curve appears less random and closer to the actual SOC line. The SOC estimation errors MAE, MSE, and RMSE, computational time, and MaxErrors are shown in Fig. 11-13. The CNN-SBLSTM algorithm performed well for SOC estimation across all driving cycles. Especially, the lowest MAE, MSE, and RMSE under the LA92 driving cycle were 0.02538, 0.0009, and 0.0301 demonstrating superior accuracy in SOC estimation under these particular driving environments. The US06 exhibits estimated errors with slightly lower MAE, MSE, and RMSE, 0.0237, 0.0012 0.0348 respectively than the LA92. The HWFET exhibits estimated errors with slightly lower MAE, MSE, and RMSE, 0.0355, 0.0022, and 0.0473 respectively. Although the MAE, MSE, and RMSE errors in the UDDS driving cycle are significantly larger than those observed in the other cycles, the overall performance remains quite accurate. The computational time (second) of UDDS, LA92, US06, and HWFET are 56.113, 23.0830, 13.5310, 43.3010 respectively as shown in Fig. 12. The MaxErors are obtained under UDDS, LA92, US06, and HWFET are 0.1854, 0.0964, 0.1191, 0.1428 respectively as shown Fig. 13. These outcomes demonstrate that the proposed network is highly accurate, adaptable, and able to function effectively across a variety of different working environments and computationally efficient.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:12:06 UTC from IEEE Xplore. Restrictions apply.

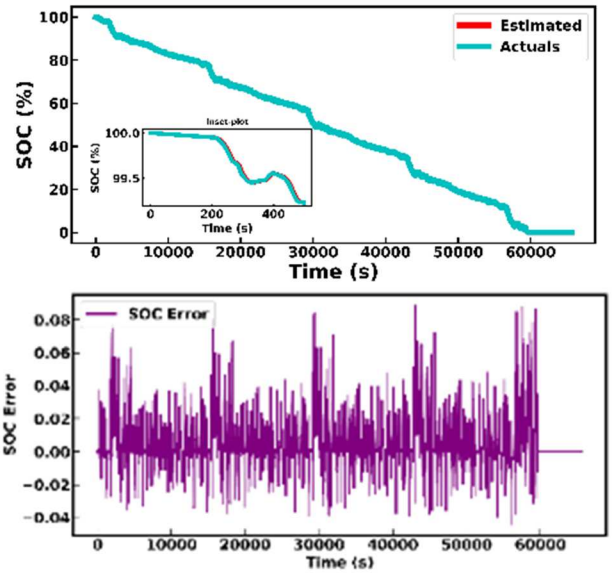


Fig. 7: Estimation performance of CNN-SBLSTM for UDDS at 25◦C.

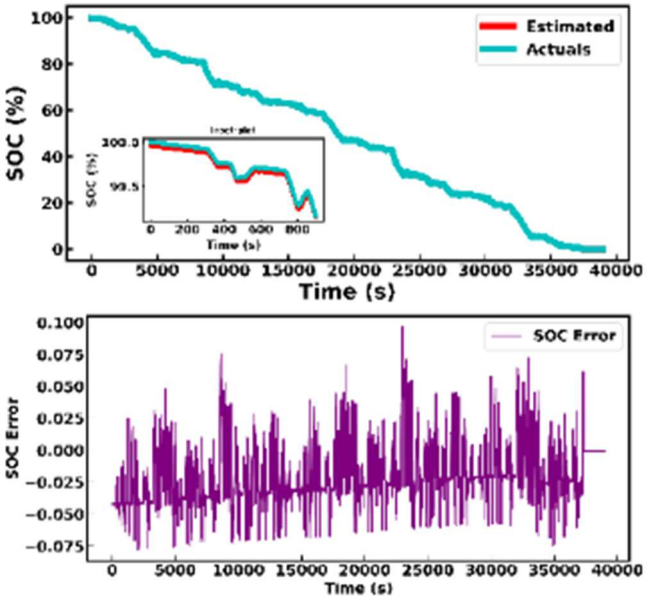


Fig. 8: Estimation performance of CNN-SBLSTM for LA92 at 25◦C.

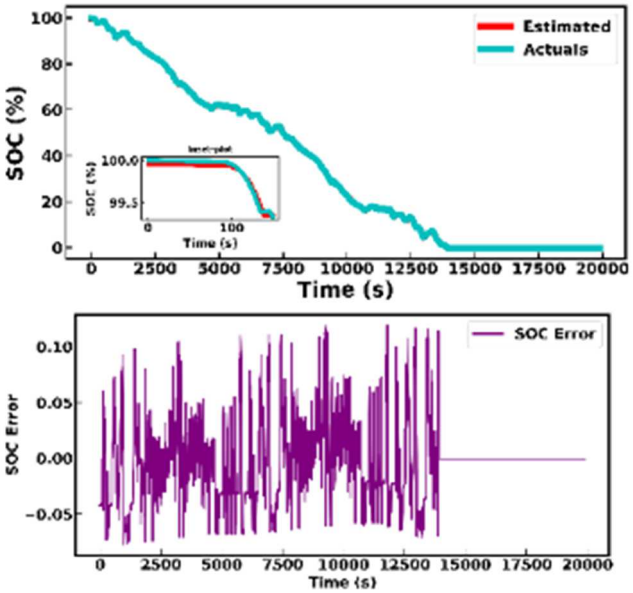


Fig. 9: Estimation performance of CNN-SBLSTM for USS06 at 25◦C.

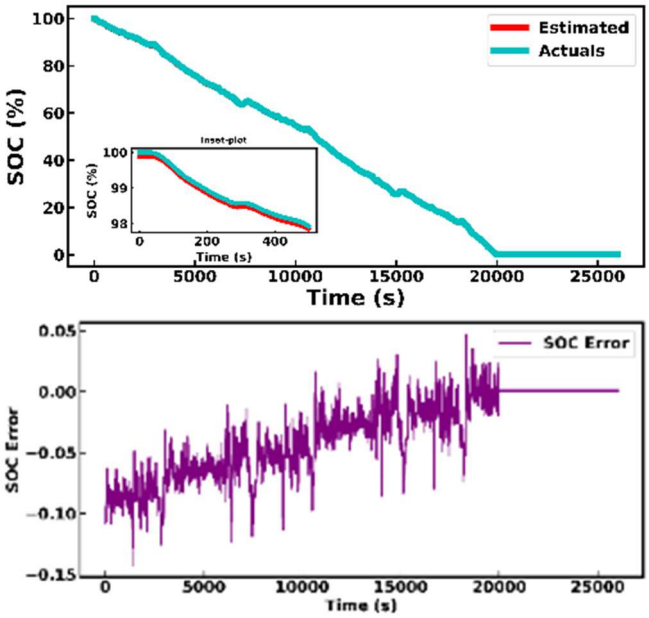


Fig. 10: Estimation performance of CNN-SBLSTM for HWFET at 25◦C.

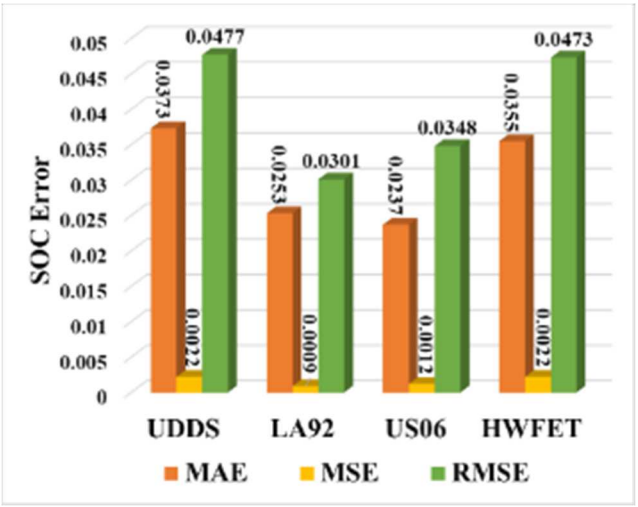


Fig. 11: SOC estimation error with CNN-SBLSTM at 25◦C under all drive cycles.

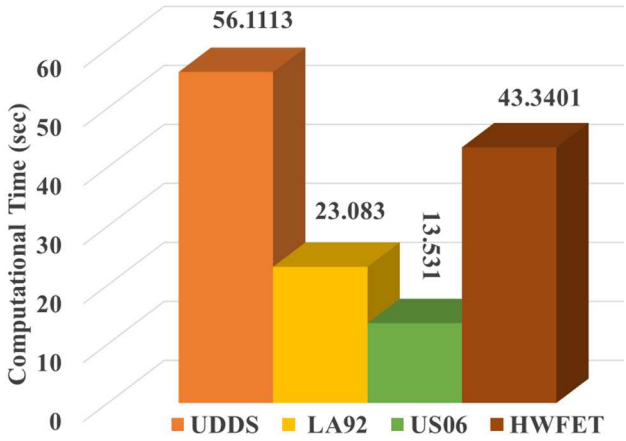


Fig. 12. Computational time with CNN-SBLSTM at 25◦C under all drive cycles.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:12:06 UTC from IEEE Xplore. Restrictions apply.

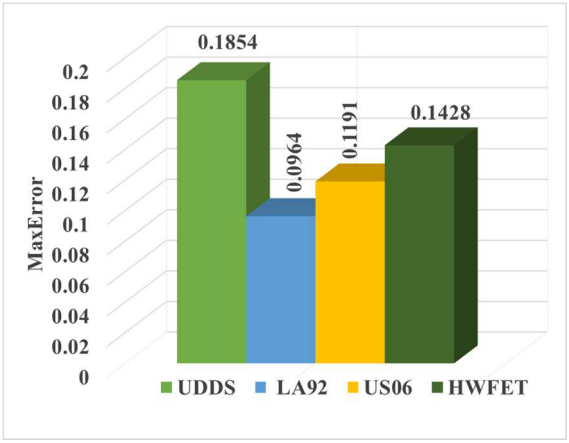


Fig. 13. MaxErrors with CNN-SBLSTM at 25◦C under all drive cycles.

IV. CONCLUSION

The main contribution of this work was the development of intelligent SOC estimations of Li-ion batteries by applying the CNN-SBLSTM approach, which improved performance, and robustness, increased estimation speed, and decreased errors under various EV driving cycles. The developed approach adaptability is investigated with four EV driving profiles at various temperatures. This study showed that the proposed approach on the LA92 cycle was shown to be capable of more precisely estimating the SOC of the UDDS, US06, and HWFET drive cycle. The SOC accuracy has increased significantly, with RMSE 0.0301% in the LA92 cycle and 0.0348% in the US06 cycle. Furthermore, the computational time is achieved at 23.083 and 13.531 seconds. The proposed model outperforms in all driving cycles. The RMSE and maximum SOC error are both estimated to be less than 0.1% and 0.2%, respectively. The state of health (SOH) for lithium-ion batteries including the effects of aging will be estimated in future studies. Furthermore, the studies will examine the performance of the proposed model with that of other machine-learning approaches.

REFERENCES

[1] A. K. Srivastava, "Air pollution: Facts, causes, and impacts," *Asian Atmospheric Pollution: Sources, Characteristics and Impacts*, pp. 39–54, Jan. 2022.

[2] S. Thangavel, D. Mohanraj, T. Girijaprasanna, S. Raju, C. Dhanamjayulu, and S. M. Muyeen, "A Comprehensive Review on Electric Vehicle: Battery Management System, Charging Station, Traction Motors," *IEEE Access*, vol. 11, pp. 20994–21019, 2023.

[3] Y. Liu, R. Zhang, J. Wang, and Y. Wang, "Current and future lithium-ion battery manufacturing," *iScience*, vol. 24, no. 4, p. 102332, Apr. 2021.

[4] Q. Shi, Z. Jiang, Z. Wang, X. Shao, and L. He, "State of charge estimation by joint approach with model-based and data-driven algorithm for lithium-ion battery," *IEEE Trans Instrum Meas*, 2022.

[5] E. Liu, G. Niu, D. Lyu, T. Yang, and B. Zhang, "Low-Cost Adaptive LS-DEKF for SOC Estimation and RDT Prediction with SFP Model," *IEEE Trans Instrum Meas*, 2023.

[6] Q. Zhu, L. Li, X. Hu, N. Xiong, and G. Di Hu, "H∞-based nonlinear observer design for state of charge estimation of Lithium-ion battery with polynomial parameters," *IEEE Trans Veh Technol*, vol. 66, no. 12, pp. 10853–10865, Dec. 2017.

[7] Z. Liu, S. Chen, B. Jing, C. Yang, J. Ji, and Z. Zhao, "Fractional variable-order calculus based state of charge estimation of Li-ion battery using dual fractional order Kalman filter," *J Energy Storage*, vol. 52, p. 104685, Aug. 2022.

[8] M. Zhang *et al.*, "A Review of SOH Prediction of Li-Ion Batteries Based on Data-Driven Algorithms," *Energies 2023, Vol. 16, Page 3167*, vol. 16, no. 7, p. 3167, Mar. 2023.

[9] C. Vidal, P. Malysz, P. Kollmeyer, and A. Emadi, "Machine Learning Applied to Electrified Vehicle Battery State of Charge and State of Health Estimation: State-of-the-Art," *IEEE Access*, vol. 8, pp. 52796–52814, 2020.

[10] Z. Wu, Q. Wang, J. X. Hu, Y. Tang, and Y. N. Zhang, "Integrating model-driven and data-driven methods for fast state estimation," *International Journal of Electrical Power & Energy Systems*, vol. 139, p. 107982, Jul. 2022.

[11] E. Chemali, P. J. Kollmeyer, M. Preindl, R. Ahmed, and A. Emadi, "Long Short-Term Memory Networks for Accurate State-of-Charge Estimation of Li-ion Batteries," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 8, pp. 6730–6739, Aug. 2018.

[12] F. Yang, W. Li, C. Li, and Q. Miao, "State-of-charge estimation of lithium-ion batteries based on gated recurrent neural network," *Energy*, vol. 175, pp. 66–75, May 2019.

[13] R. R. Ardeshiri and C. Ma, "State of Charge Estimation of Lithium-ion Battery Using Deep Convolutional Stacked Bidirectional LSTM," *IEEE International Symposium on Industrial Electronics*, vol. 2021-June, Jun. 2021.

[14] B. Zraibi, C. Okar, H. Chaoui, and M. Mansouri, "Remaining Useful Life Assessment for Lithium-Ion Batteries Using CNN-LSTM-DNN Hybrid Method," *IEEE Trans Veh Technol*, vol. 70, no. 5, pp. 4252–4261, May 2021.

[15] Q. Wang, M. Ye, M. Wei, and G. Lian, "Optimized deep neural network enabled low-cost state of charge estimation for different kinds of lithium-ion batteries under dynamic load conditions," *Int J Energy Res*, vol. 46, no. 15, pp. 22946–22959, Dec. 2022.

[16] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, no. 4, pp. 611–629, Aug. 2018.

[17] H. Chaoui and C. C. Ibe-Ekeocha, "State of Charge and State of Health Estimation for Lithium Batteries Using Recurrent Neural Networks," *IEEE Trans Veh Technol*, vol. 66, no. 10, pp. 8773–8783, Oct. 2017.

[18] P. Eleftheriadis, S. Leva, and E. Ogliari, "Bayesian Hyperparameter Optimization of Stacked Long Short-Term Memory Neural Network for the State of Charge Estimation," *SyNERGY MED 2022 - 2nd International Conference on Energy Transition in the Mediterranean Area, Proceedings*, 2022.

[19] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[20] F. U. M. Ullah, A. Ullah, I. U. Haq, S. Rho, and S. W. Baik, "Short-Term Prediction of Residential Power Energy Consumption via CNN and Multi-Layer Bi-Directional LSTM Networks," *IEEE Access*, vol. 8, pp. 123369–123380, 2020.

[21] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values," *Transp Res Part C Emerg Technol*, vol. 118, p. 102674, Sep. 2020.

[22] D. N. T. How, M. A. Hannan, M. S. H. Lipu, K. S. M. Sahari, P. J. Ker, and K. M. Muttaqi, "State-of-Charge Estimation of Li-Ion Battery in Electric Vehicles: A Deep Neural Network Approach," *IEEE Trans Ind Appl*, vol. 56, no. 5, pp. 5565–5574, Sep. 2020.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:12:06 UTC from IEEE Xplore. Restrictions apply.

International Journal of Hydrogen Energy 59 (2024) 1080–1093

Contents lists available at ScienceDirect

International Journal of Hydrogen Energy

journal homepage: www.elsevier.com/locate/he

Advancements in composite cathodes for intermediate-temperature solid oxide fuel cells: A comprehensive review

Anil Kumar Yadav [a], Shailendra Sinha [a], Anil Kumar [b,c,*]

[a] Department of Mechanical Engineering, Institute of Engineering & Technology, Lucknow, 226021, India
[b] Department of Mechanical Engineering, Delhi Technological University, Delhi, 110042, India
[c] Centre for Energy and Environment, Delhi Technological University, Delhi, 110042, India

ARTICLE INFO

Handling Editor: Fanglin F. Chen

Keywords:
Solid oxide fuel cell
Electrochemical performance
Composite material
Oxygen reduction reaction

ABSTRACT

Fuel cells have substantial development potential for commercial energy generation applications. Adding an ionic conducting electrolyte to a cathode to make a composite cathode lowers the Thermal Expansion Coefficient (TEC), raises the Triple Phase Boundary (TPB) area, and enhances the electrochemical performance of Solid Oxide Fuel Cells (SOFCs). In this review, the article presents the basic principles, criteria, and composite cathode development of SOFCs that operate SOFCs in the intermediate temperature ranges (600–800 °C). These composite cathodes improve the inter-component compatibility, Oxygen Reduction Reaction (ORR), and characteristics of mixed electronic-ionic conductors. In comparison with a single-phase cathode, composite material can reduce Polarization Resistance (R_p) by about 30–40% and improve maximum power density (MPD) by about 30–50%. PBCO-40SDC and SBCN-20SDC composites exhibit enhanced stability and maintain their electrochemical performance with polarization resistances of 0.008 Ω cm^2 and 0.079 Ω cm^2 when operating at temperature of 700 °C.

1. Introduction

The greatest challenge of the present scenario is clean energy production. Fossil fuels for energy generation and transportation create the bulk of pollutants and harmful gases that contribute to climate change [1,2]. As a result, drastic reductions in these emissions are required to avoid irreversible harm to our environment [3,4]. Reduced greenhouse gas emissions are a significant issue since the present energy production system primarily depends on fossil fuels, coal, and oil [5,6]. Fuel cell-based power-generating systems are gaining much attention in research and industry because they can supply consumers with high-potential, long-lasting clean energy [7,8]. This technology benefits from high energy conversion, clean energy production, high-efficiency energy generation, and fuel flexibility [9].

The critical components of SOFC are an electrolyte and two distinct electrodes (anode and cathode) [10]. The two porous electrodes are separated by a ceramic electrolyte, as illustrated in Fig. 1a. In SOFCs, the working process begins by reducing oxygen in the cathode to create the oxide ion. The solid electrolyte then transports the oxide ion to the anode, completing the fuel oxidation process and yielding an electron. Finally, the electron passes through the external load to finish the process of generating electrical energy [11]. Fig. 1b and c depicts geometrical configurations, i.e. planar and tubular designs. Presently, researchers are working on decreasing the operating temperature of SOFC [12]. It will allow it to be used more efficiently and reduce CO_2 emissions [13]. To reduce the working temperatures of SOFCs, it is essential to provide more active electrodes [14,15]. According to the Arrhenius conduction equation, the electrolyte ionic conductivity falls exponentially as the functioning temperature decreases [16]. Despite the low-temperature SOFC's enjoyable success and technological benefits, low-temperature operation increases electrolyte ohmic loss and electrode polarization losses, resulting in decreased overall cell performance and energy conversion efficiency [17,18]. SOFC researchers and manufacturers have made multiple attempts to reduce the working temperature to intermediate ranges (600–800 °C) and even lower temperatures (below 600 °C) to overcome these obstacles [19,20]. According to several studies, for optimal performance of SOFCs, the functioning temperature lies between 400 °C and 700 °C. The cells' cathodic polarization resistance should be within acceptable limits at this temperature range [21,22].

In SOFC, cathode materials form the interface between the electrode-electrolyte-oxidant, i.e. triple phase boundary, and reduce TEC of the

* Corresponding author. Department of Mechanical Engineering, Delhi Technological University, Delhi, 110042, India.
E-mail address: anilkumar76@dtu.ac.in (A. Kumar).

https://doi.org/10.1016/j.ijhydene.2024.02.124
Received 23 December 2023; Received in revised form 7 February 2024; Accepted 9 February 2024
Available online 14 February 2024
0360-3199/© 2024 Hydrogen Energy Publications LLC. Published by Elsevier Ltd. All rights reserved.

material [23]. The thermal expansion analysis indicates that a higher electrolyte content significantly diminishes the TEC. Introducing a moderate amount of ionic-conducting electrolyte into the cathode material typically enhances oxygen diffusion rates and facilitates the transfer of oxygen ions at the electrolyte/electrode interface. However, an excess of the electrolyte phase in the composite cathode results in a notable decrease in TPB areas and an increase in network interruptions, encompassing both the cathode and electrolyte [24]. The cathode materials must possess a high level of electrochemical activity in order to effectively catalyze the oxygen reduction reaction while also facilitating the rapid diffusion of oxygen anions across the cathode [25]. By adding an electrolyte to the cathode, for instance, the TPB's size is expanded, enhancing electrochemical performance while simultaneously decreasing TEC [26]. MIECs (mixed ionic-electronic conductors) are a composite material that was created primarily to improve the electrocatalytic performance of SOFCs [27]. Penetrating the oxygen reduction sites into the bulk of the electrodes in MIEC cathodes can expand the oxygen-reducing region from the TPBs to the majority of the cathode, reducing cathode polarization losses [28]. The recovery of Pb from discarded lead-acid batteries contributes to the preservation of energy resources and the mitigation of environmental pollution. At 600 °C, the maximum open circuit voltage (0.95 V) and the maximum power density (439 mW cm^{-2}) have been achieved using hydrogen as working fuel. Lower value of activation energy (0.36 eV) of $Pb_{0.3}Fe_{0.2}Co_{0.5}O_{4-\delta}$ confirms that it is an effective material for converting electrochemically hydrogen fuel into valuable electricity [29].

Different strategies for improving the functionality of the cathode materials are doping methods [30,31], nanostructuring [32], composites [33], etc. The need for a composite cathode permits the reaction zone to expand into the electrode from the electrode/electrolyte interface [34]. $(La_{0.6}Sr_{0.4})_{0.95}Co_{0.2}Fe_{0.8}O_{3-\delta}$ (LSCF) has been extensively studied for an extended period. However, its limited ORR activity and

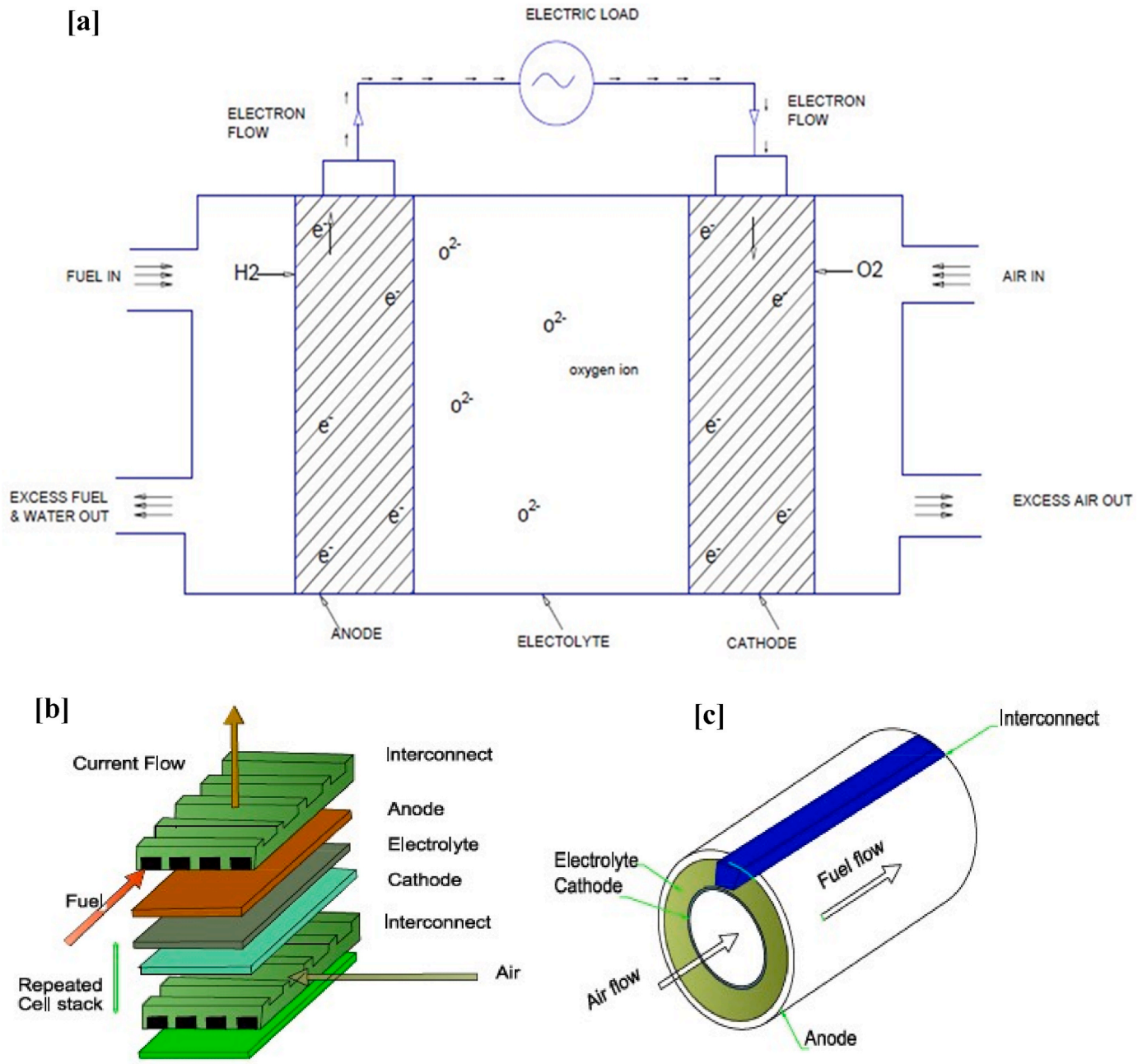


Fig. 1. (a) SOFC principle of operation; (b) Planar design; (c) Tubular design.

inherent performance deterioration remain significant hurdles to its widespread use in SOFCs with oxygen-ion conduction. A straightforward infiltration technique has successfully deposited a Pr_6O_{11}/NiO composite on LSCF. In comparison to LSCF, polarization resistance of the improved LSCF cathode was roughly 30% lower [35]. Yurchenko et al. [36] investigated the performance of $PrBaCo_2O_{6-\delta}$ (PBCO)–$Ca_3Co_4O_{9\pm\delta}$ (CCO) composite cathode. With increasing CCO content in composite materials, TEC declines perceptibly with a minimal electrical conductivity decline. This composite material offers 70% and 97% less area-specific resistance than pure PBCO and CCO, respectively. Composite cathode $SmBa_{0.5}Sr_{0.5}Co_2O_{5+\delta}$ (SBSCO)–$Ce_{0.9}Gd_{0.1}O_2$ exhibits about 86% less ASR than single-phase SBSCO cathode material at 600 °C [37]. Xia et al. [38] explored an innovative approach that combines doping and surface modification to enhance the catalytic activity and CO_2 resistance of the $Pr_{0.4}Sr_{0.6}Co_{0.2}Fe_{0.8}O_{3-\delta}$ cathode. The polarization resistance of $PSCFN_{0.1}$ at 800 °C recorded a 0.025 Ω cm^2 value, indicating a significant 32.4% reduction compared to PSCF. Following this, a solution of $Pr(NO_3)_3$ was applied to the surface of $PSCFN_{0.1}$ using a spin-coating technique at 800 °C, resulting in a polarization resistance of only 0.016 Ω cm^2 for PrO_2-$PSCFN_{0.1}$, a remarkable 56.8% reduction compared to PSCF.

Duan et al. [39] investigated $CuMn_{1.5}Ni_{0.5}O_4$ (CMN)-GDC composite cathode performance at 800 °C. It gives 0.047 Ω cm^2 polarization resistance and 1342 mW cm^{-2} maximum power density (MPD). Authors recommended that after 10 h, cells operate stably with no microstructure problems. In order to deal with degradation of cathode material due to the poisoning contaminants in ambient air, Li et al. [40] studied the performance of $SrCo_{0.8}Nb_{0.1}Ta_{0.1}O_{3-\delta}$ (SCNT)-La_2NiO_4 (LNO). Composite material offers 33% less polarization resistance at 600 °C and also improves ORR activity as compared to single phase SCNT. Xu et al. [41] prepared $(La_{0.85}Sr_{0.15})_{0.9}MnO_{3-\delta}$-$Sm_{0.2}Ce_{0.8}O_{1.9}$ using a two-step fabrication method involving screen-printing and ion-impregnation. At 700 °C, the LSM-SDC electrode with impregnated LSM exhibited a significantly lower total resistance of 0.35 Ω cm^2, contrasting with the 0.94 Ω cm^2 resistance observed in a conventional electrode. In the temperature range of 550 °C–700 °C, the $SrNb_{0.1}Co_{0.9}O_{3-\delta}$-$20Sm_{0.85}Zn_{0.15}MnO_3$ cathode symmetric cell demonstrated a low Rp of only 0.3654–0.012 Ω cm^2. Furthermore, the SNC-20SZM-based fuel cell utilizing anode support exhibited a notable 44–49% increase in MPD (0.62–1.22 W cm^{-2}) [42]. As mentioned in the above literature, composite materials could play a key role in lowering electrode potentials by providing a larger effective surface area for electrochemical reactions.

In the present work current research areas of SOFC composite cathodes have been explored. Several recent and important composite method developments i.e. exsolution of catalytically active nanoparticles, one-pot method, and in-situ techniques have been discussed. Developing novel cathode materials and microstructure optimization for low temperature (LT) and IT-SOFC. Comparison of nine selected materials on the basis of polarization resistance, peak power density, thermal expansion coefficient, and electrical conductivity has been discussed for IT-SOFC.

2. Overview of cathode reaction mechanism

The oxygen reduction process (ORR) converts gaseous O_2 into solid oxygen ions (O^{2-}) at the cathode [43]. The exchange mechanism involves oxygen adsorption, oxygen dissociation, transportation of oxygen ions through cathode bulk or along the surface, followed by incorporation of oxygen ions into the electrolyte [44]. Fig. 2 shows that ORR mechanism is possible along the triple phase boundary or by bulk oxide ion transport. The triple phase boundary (TPB) is the precise interface where the gaseous phase of oxygen, the oxygen ion conductor, and the electronic conductor meet [45]. On the cathode surface, ORR mechanism is characterized by equation (1) [46].

$$\frac{1}{2}O_2(gas) + 2e^-(cathode) \rightarrow O^{2-}(electrolyte) \quad [1]$$

The cathode plays a significant role in determining the overall performance of a solid oxide fuel cell, contributing around 50% of the total polarization resistance [47]. Electrode polarization typically originates at the cathode due to the larger activation energy of ORR compared to the hydrogen oxidation process at the anode [48]. A major challenge in reducing the operating temperature of SOFCs in the 400–600 °C region

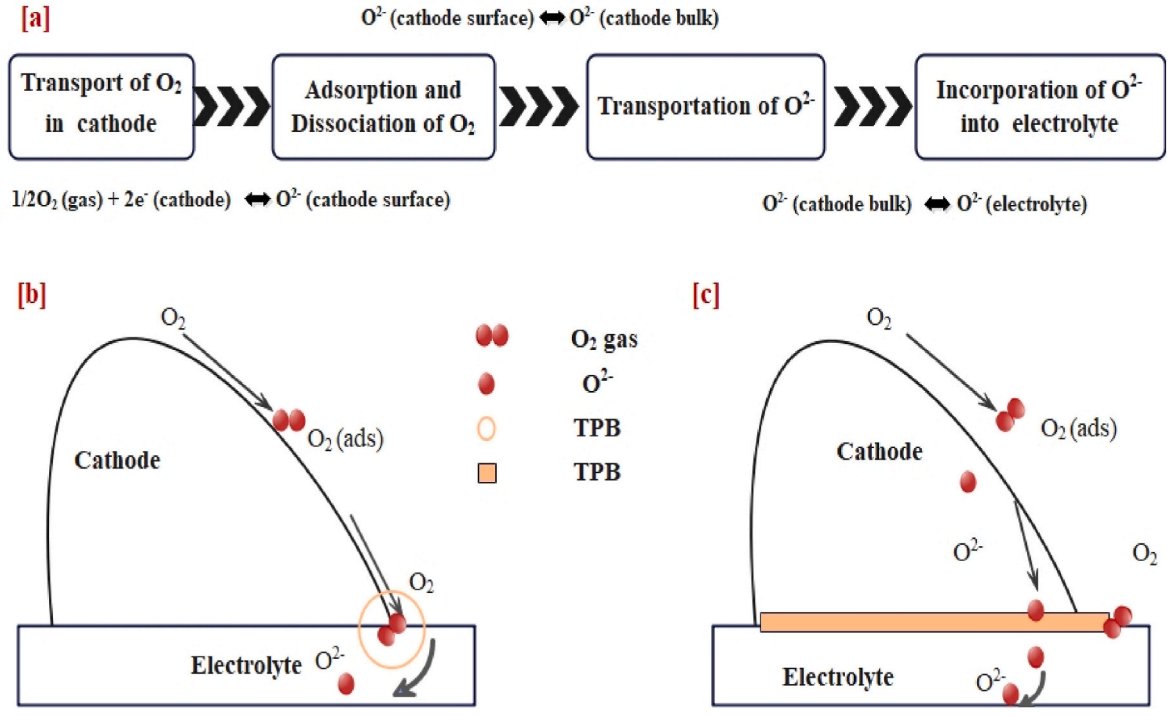


Fig. 2. (a) ORR steps in the cathode; (b) ORR along the TPB; (c) ORR along bulk transport of oxide ions.

is cathode polarization, which is caused by the large activation energy required for the ORR mechanism [49,50]. Finding cathode materials that show rapid ORR even at moderate to low temperatures is, therefore, of the utmost importance [51]. Cathodes with a compact, thin film arrangement show improved ORR along the TPB, which is influenced by the film's thickness and porosity [52]. Raising the material's porosity facilitates gas quick diffusion, which in turn decreases concentration polarization and improves cathode-oxygen gas interaction [53]. In this regard, the creation of chemically stable composite cathodes with improved performance has received a lot of attention during the last several decades. Reducing the electrode polarization entails increasing the number of active sites, or TPB [54].

Composite cathodes, which include additional oxygen ion conductors added to the original electrode materials, are the most common way to expand TPB area. Adding oxygen ion conductors raises the oxygen vacancy concentration, which in turn increases the number of oxygen ion transport channels, which in turn improves oxygen diffusion by extending the TPB to the whole electrode bulk instead of just the electrode-electrolyte interface [55]. To improve the ORR's electro-catalytic performance at low temperatures (>650 °C), it is possible to alter the oxygen stoichiometry and other physiochemical characteristics by adding mixed ionic and electronic conduction (MIEC) to the cathode material [56]. The oxygen vacancy transit in porous MIEC electrodes along the ORR across the active electrode thickness makes kinetic identification more difficult. This is what causes the typical hedgehog shape in the Nyquist plot. The MIEC electrode is better at conducting ions because it has less oxygen pressure and oxygen vacancies because of the cathode bias [57]. The cathodic behavior of a brownmillerite-structured composite of $Ca_2Fe_2O_5$–$Ca_2Co_2O_5$ was investigated by Pichandi et al. [58]. The MIEC properties allowed for the expansion of the triple-phase boundary region, which in turn improved the ORR and its kinetics. The fact that the symmetric cell and the half-cell both had low activation energies reported it. Aziz et al. [59] fabricated composite cathodes consisting of $SrFe_{0.9}Ti_{0.1}O_{3-\delta}$ and $Ce_{0.8}Sm_{0.2}O_{1.9}$ with varying layers. By extending the active area for the reaction up to the threshold point, the performance of the cathode could be enhanced by adding more layers to the cathode.

Another tough problem is that as the working temperature drops, the catalytic activity for ORR in the SOFC cathode decreases. Composite cathodes often use more than one material to improve their catalytic activity, but it's still hard to get the best rates. Yu et al. [60] investigated the performance of $PrBaCo_2O_{5+\delta}$-GDC cathode at 800 °C. Cathode gives 0.996 W cm^{-2} maximum power density and good stability in 216 h. Reducing polarization resistance at high and intermediate frequencies was largely responsible for the material's outstanding performance. This reduction leads to an improved transfer process of O^{2-}, oxygen (O_2) surface exchange, and charge transfer. Liu et al. [61] said that the heterostructure's best ORR activity was reached when the amount of Pr_6O_{11} particles incorporated reached about 10% of the $Pr_2NiO_{4+\delta}$. The oxygen surface exchange coefficient of 2-Pr_6O_{11}–$Pr_2NiO_{4+\delta}$ was 1.45 times greater than that of $Pr_2NiO_{4+\delta}$ at 750 °C. The surface exchange coefficients (K_{ef}) and diffusion coefficients (D_{ef}) of LSCF-$La_2Ce_2O_{7+\delta}$ (LCO) were 8×10^{-3} cm s^{-1} and 4×10^{-4} cm^2 s^{-1} at 700 °C. It was higher than LSCF- $Sm_{0.2}Ce_{0.8}O_2$ (SCO) and LSCF-$BaZr_{0.1}Ce_{0.7}Y_{0.2}O_{3-\delta}$ (BZCY), suggesting LCO enhances surface adsorption ability, which plays an important role in ORR [62].

3. Research progress in composite cathode materials of SOFCs

By modifying the cathode microstructure in various ways and adding electrolyte particles or metal catalysts to the oxide cathode material, numerous composite cathode materials have been suggested and studied in the majority of the literature. Few researchers have created co-doping composite electrode materials that do not react with electrolytes [63]. Yang et al. [64] developed a new perovskite oxide, $Sm_{0.8}Ca_{0.2}Ba_{1-x}Ca_xCo_2O_{5+\delta}$ (x = 0.1 and 0.2), using calcium co-doping. At 800 °C, the x = 0.1 cathode has a specific resistance of 0.022 Ω cm^2 and a maximum power density of 885 mW cm^{-2}. The x = 0.1 double perovskite sample exhibits good chemical compatibility and electro-chemical characteristics and considerably reduces TEC, making it a viable IT-SOFC material. The presence of the electron conducting phase in composite cathodes make for a rather circuitous diffusion channel for oxide ions; however, this could be circumvented by the use of infiltration and impregnation procedures using microstructures that are well-turned [65]. Since energy conversion technologies are in high demand, it is crucial to find novel electrolytes that exhibit excellent ionic conductivity at low temperatures [66]. More study into cathode procedures, coating and infiltration techniques, doping approaches, nanostructuring, and microstructure control, as well as degradation mechanisms, is needed for composite cathode material advancement in SOFC applications [67].

The utilization of nanostructuring technology provides several benefits owing to its elevated surface to volume ratio and heightened reactivity [68,69]. The infiltration exhibits distinctive features, including reduced dimensions and elevated catalytic efficacy [70]. Nano-structured electrodes fabricated using infiltration or impregnation at low temperatures offer enhanced activity and a wider range of material choices. Nano-catalysts that conduct electricity are mixed with single-phase ionic backbones, and nanoparticles that conduct electricity are mixed with backbones that conduct electricity [71]. Fan et al. [72] looked at SSC-GDC composite cathodes with nanofiber structures. These were created by adding a precursor solution of $Gd_{0.2}Ce_{0.8}O_{1.9}$ into a porous $Sm_{0.5}Sr_{0.5}CoO_{3-\delta}$ nanofiber scaffold. By combining the nanofiber-structured SSC-GDC cathode with a 650 μm-thick $La_{0.8}Sr_{0.2}Ga_{0.8}Mg_{0.2}O_{3-\delta}$ electrolyte, the maximum power density of a cell at 800 °C was determined to be 0.134 W cm^{-2}. Impregnation method can be used to attach a catalytic material to the surface of another material to produce a new composite electrode [73]. The impregnation composite refers to the combination of particles and the framework, enhancing the attachment of other-phase materials to the electrolyte skeleton or cathode framework. This process effectively increases the specific surface area of the cathode skeleton, extends the length of TPB, and augments the number of reaction activation sites, ultimately improving the electrochemical performance of the cathode materials [74]. In a study by Gao et al. [75], nano-scale $SrFe_{0.75}Mo_{0.25}O_{3-\delta}$ was impregnated onto a $La_{0.9}Sr_{0.1}Ga_{0.8}Mg_{0.2}O_{3-\delta}$ composite material. The results indicate that at 700 °C, the composite cathode exhibited a polarization resistance of 0.25 Ω cm^2 and a peak power density of 521 mW-cm^{-2}. To improve the composite cathode's surface electron transfer rate and get around the fact that PBCO does not work well with GDC electrolyte when it comes to matching thermal expansion, it can be soaked into the surface of the LSCF skeleton. The electrode's oxygen exchange rate was also accelerated [76]. Cathodes for solid oxide fuel cells operating at intermediate temperatures were made with high performance (La,Sr)MnO_3 by ion impregnation of oxygen ion conducting Gd-doped CeO_2 oxide. A range of 100–200 nm in size made up the impregnated GDC phase. A polarization resistance of 0.21 Ω cm^2 was measured at 700 °C for the 5.8 mg/cm^2 GDC-impregnated LSM electrode, which is 56 times lower than the basal LSM at the same temperature [77].

In situ exsolution is a commonly used method for producing nanoparticles, especially in perovskite materials [78]. In contrast to the infiltration method, this one lets move and dissolve metallic nanoparticles without worrying about how the scaffold is shaped. By releasing metals that are easily reduced from their oxide lattice structures, this process enables them to move toward the surface and, when heated, form nanoparticles. The exsolved nanoparticles stay bound to the surface of the first phase, which is a distinctive feature of this method that reduces conditions [79]. Exsolution is a sequential process involving diffusion, reduction, nucleation, and growth. The reduced metal initiates nucleation on the surface, expanding as the treatment time extends. The particle size and morphology are significantly

influenced by factors such as the upper limit content of active metal ions, segregation energy of metal and oxygen ions, and external environmental parameters like treatment temperature, reducing gas atmosphere, and treatment duration [80]. In a study by Peng et al. [81], $CaCuO_2$ was exsolved with $SmBa_{0.5}Ca_{0.5}CoCuO_{5-\delta}$. The results indicate that the exsolved cathode with GDC electrolyte exhibited a thermal expansion coefficient of 14.3×10^{-6} K^{-1} and a peak power density of 1230 mW-cm^{-2} at 800 °C.

The performance of a self-assembled nanocomposite of $La_{0.6}Sr_{0.4}FeO_{3-\delta}$-$La_{1.2}Sr_{0.8}NiO_{4+\delta}$ was investigated by Yang et al. [82]. Because the two phases are evenly distributed at the nanoscale in the one-pot synthesized composite cathode, it exhibits an enhanced synergistic effect. At 700 °C in oxygen, the cathode demonstrates the lowest polarization resistance of 0.055 Ω cm^2, while in air, it shows the lowest resistance of 0.095 Ω cm^2. There has been a lot of recent focus on creating new composite cathodes with a core-shell structure [83]. By adding a highly-active surface layer, this novel microstructure can boost electrochemical performance; this layer often has a greater ORR performance than the bare core [84]. By creating a conformal surface coating that effectively suppresses the unwanted thermal growth of grains and chemical interactions with impurities, the stability of the cathode can be enhanced. The porous heterostructured $Nd_{0.8}Sr_{1.2}CoO_{4\pm\delta}$-$Nd_{0.5}Sr_{0.5}CoO_{3-\delta}$ exhibits improved electrochemical properties when contrasted with single-phase materials. These include enhanced conductivity, kinetics for surface oxygen exchange that are twenty times faster, and a relatively lower ASR of 0.065 Ω cm^2 at 800 °C [54]. The SFT-SDC heterostructure has an MPD of 920 mW cm^{-2} at 520 °C, owing to its strong ionic conductivity >0.1 S cm^{-1}, which is one order of magnitude greater than that of bulk SDC [85]. Effect of different modification strategies on the performance of composite cathode materials is depicted in Table 1.

4. Essential properties of composite cathode materials

Even though SOFCs' working concept is straightforward, selecting materials for the various components is tremendously difficult [107,108]. The electrode materials must possess high catalytic activity to facilitate the intended chemical and electrochemical reactions [109,110]. The microstructure of the electrodes must have a considerable number of active reaction sites [111,112]. Every component of the stack must be chemically stable, taking into consideration variations in gas composition due to fuel and oxygen utilization, fuel supply issues, and stack leaks [113,114]. Thermal expansion, chemical compatibility, and surface adhesion must all be met for the materials to be compatible [115,116]. Different synthesis techniques, as shown in Table 2, such as sol-gel [117,118], co-precipitation [52,119], solid-state reaction [120,121], glycine nitrate process [122,123], EDTA-citrate method [124,125], etc. are widely used to synthesize composite cathodes. Sintering temperature influences oxidation and electrochemical coating properties. It is difficult to avoid interdiffusion when SOFCs are being sinterized and used because of the high temperatures and chemical potential gradients [126]. Composite cathode LSCF-SCDC possesses a maximum power density of 543 mW cm^{-2} when it was sintered at 600 °C [127].

4.1. Electrical and thermal properties

The available literature shows composite cathode materials' electrical and thermal properties, as shown in Fig. 3a–e and Table 3. Fig. 3a illustrates the electrical conductivity of nine materials primarily used in the composite cathode. High electrical conductivity is necessary to

Table 1
Effect of different strategies on the performance of composite cathode.

Base material	Strategy	Benefits	Ref.
$Sr_{0.74}La_{0.26}CoO_{3-\delta}$-$SrZr_{0.79}Co_{0.2}O_{3-\delta}$	One-pot in situ self-assembly	• Excellent performance (MPD of 2668 mW cm^{-2} at 750 °C) and hierarchical structure's active surface enhance ORR activity.	[86]
$NdBaCo_2O_{5+\delta}$-$xCe_{0.9}Gd_{0.1}O_{1.95}$	In-situ sol mixing	• After being sintered at 1100 for 12 h, composite cathodes do not undergo any chemical reactions. • CGO nanoparticles increase surface oxygen dissociation.	[87]
$PrBaCo_2O_{5+\delta}$-$Gd_{0.1}Ce_{0.9}O_{2-\delta}$	One-pot synthesis	• PBC and GDC particles produce a coherent interface structure that reduces lattice thermal expansion.	[88]
PrLaNiO4-5LSBC	Mechanical-physical composite	• At 800 °C, composite cathode exhibits 606 mW cm^{-2} MPD and 0.202 Ω cm^2 ASR	[89]
$La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_{3-\delta}$-$Er_{0.4}Bi_{1.6}O_3$	Infiltration using organic solvents	• Rp reduces from 0.27 to 0.11 Ω cm^2 and activation energy from 100.28 to 97.15 kJ/mol at 700 °C	[90]
$La_{0.1}Sr_xCa_{0.9-x}MnO_{3-\delta}$-$Sm_{0.2}Ce_{0.8}O_{1.9}$	Catalyst coating	• MPD of 800 mW cm^{-2} at 550 °C	[91]
$La_{0.9}Sr_{0.1}CoO_3$	Doping (Fe,Ni,Zn)	• Fe or Ni doping into cathode lowers surface oxygen and ORR catalytic activity. • Fe doped cathode gives MPD of 239.1 mW cm^{-2} at 700 °C	[92]
$La_2NiO_{4+\delta}$	Doping (Co)	• Improves lanthanum nickelate electrochemistry	[93]
$La_{0.8}Sr_{0.2}MnO_{3-\delta}$-$Ce_{0.9}Gd_{0.1}O_{1.95}$	Infiltration using spray pyrolysis deposition	• A strong ionic conductivity in the backbone layer leads to a decrease in Rp	[94]
$LiNi_{0.8}Co_{0.15}Al_{0.05}O_{2-\delta}$	Fluorine doping	• Improves oxygen ionic conductivity and decreases ASR and activation energy on cell	[95]
$La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_{3-\delta}$-$Ce_{0.8}Sm_{0.2}O_{1.9}$	One step microwave self assisted	• Lowered total Rp compared to literature reports of identical cathodes synthesized by different methods	[96]
$La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_{3-\delta}$-SDC	Rotary evaporation technique	• Microstructure, crystal structure, and purity of LSCF cathode powders are closely related to properties of composite	[97]
LSCF-GDC	In situ composite	• Lower Rp as compared to mechanical-physical composite	[98]
$PrBa_{0.8}Ca_{0.2}Co_2O_{5+\delta}$-$Gd_{0.2}Ce_{0.8}O_{1.9}$	Nanostructuring	• MPD of 2080 mW cm^{-2} at 750 °C	[99]
$(La_{0.2}Nd_{0.2}Sm_{0.2}Sr_{0.2}Ba_{0.2})Co_{0.2}Fe_{0.8}O_3$-SDC	Impregnation	• Rp of nano-size cathode material was reduced by 50.8% as compared to sintered composite	[100]
$BaCo_{0.4}Fe_{0.4}Zr_{0.1}Y_{0.1}O_{3-\delta}$ $Ce_{0.8}Gd_{0.2}O_{1.9}$	Infiltration	• Rp of composite cathode with 18 mg/cm^2 PrO_x nano-particle was 25.3% of composite cathode without PrO_x	[101]
$(La_{0.6}Sr_{0.4})_{0.95}Co_{0.2}Fe_{0.8}O_{3-\delta}$	Infiltration	• $(La_{0.6}Sr_{0.4})_{0.95}Co_{0.2}Fe_{0.8}O_{3-\delta}$ with a composite of 5 μL Pr_6O_{11} and NiO offers 30% less Rp than the base cathode	[35]
$La_{0.8}Sr_{0.2}Co_{0.2}Fe_{0.8}O_{3-\delta}$	Coating layer	• Composite cathode shows no sign of degradation for 200 h of continuous use	[102]
$Pr_2NiO_{4+\delta}$	Impregnation	• ASR of $Pr_2NiO_{4+\delta}$ -$La_{0.6}Sr_{0.4}CoO_{3-\delta}$ (0.024 Ω cm^2) at 700 °C was 83.4% lower than that of base cathode (0.145 Ω cm^2)	[103]
$La_{0.8}Sr_{0.2}FeO_{3-\delta}$-$Gd_{0.2}Ce_{0.8}O_{2-\delta}$	Nanostructuring	• With 9.12 wt% $BaCO_3$ modified in based composite, the area specific resistance was 2.2 times less than that of base material	[104]
Co_3O_4-GDC	Nano-structuring	• Excellent durability under a galvanostatic load of 1 A/cm^2 and 800 °C for 300 h	[105]
$LaNi_{0.6}Fe_{0.4}O_{3-\delta}$	Additive	• Electrode polarization resistance with a Bi–Cu additive in the base material was in the range of 0.60–0.67 Ω cm^2 at 600 °C.	[106]

Table 2
Cathode materials with synthesis technique and sintering temperature.

Material composition	Synthesis Process	Sintering temperature (°C)/dwell time	Morphology	Average particle distribution (nm)	Ref.
$La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_{3-\delta}-10LaSrCo_{0.2}Fe_{0.8}O_{4-\delta}$	Wet chemical coprecipitation	1000	Porous structure	200	[128]
$Sr_{0.74}La_{0.26}CoO_{3-\delta}$-$SrZr_{0.79}Co_{0.21}O_{3-\delta}$	Modified glycine-nitrate combustion	1000/2h	Nanoparticles	100	[86]
$La_{0.8}Sr_{0.2}Co_{0.2}Fe_{0.8}O_{3-\delta}$-$Gd_{0.2}Ce_{0.8}O_{1.9}$	Electrospinning, screen printing and solution infiltration	800/1h	Nanofiber	280	[129]
$Pr_{0.9}Y_{0.1}BaCo_{1.8}Ni_{0.2}O_{6-\delta}-30Ce_{0.8}Sm_{0.2}O_{1.9}$	Organo-metallic precursors	1000/1h	–	–	[130]
$Y2O_3ZrO_2$–$La_{0.8}Sr_{0.2}MnO_3$	Solution precursor plasma spraying	1050/2h	Porous structure	200	[131]
(La,Sr)CoO_3-$(La,Sr)_2CoO_4$	Induction plasma system	–	–	30	[132]
$Ba_{0.5}A_{0.5}Fe_{0.8}Zr_{0.2}O_{3-\delta}$ (A = Sr^{2+}/Sm^{3+})	Modified glycine-nitrate auto ignition	900/2h	Nanoparticles	40–100	[133]
$Ba_{0.5}Sr_{0.5}Co_{0.8}Fe_{0.2}O_{3-\delta}$-$Sm_{0.2}Ce_{0.8}O_{1.9}$	Ball-milling	550–750	Pellet morphology	114	[134]
$LaSrCoO_4$–35$Ce_{0.9}Mn_{0.1}O_2$	Citric acid-nitrate	950	Porous structure	–	[55]
$La_{0.8}Sr_{0.2}MnO_{3-\delta}$-$Ce_{0.9}Gd_{0.1}O_{1.95}$	Spray-pyrolysis deposition	1100/1h	Coated pellets	100	[94]
$La_{0.8}Sr_{0.2}Co_{0.2}Fe_{0.8}O_{3-\delta}$-$Gd_{0.2}Ce_{0.8}O_{1.9}$	Sol-gel	900/2h	Nanorods/ nanoparticles	140	[135]
CO_3O_4-GDC	Glycine nitrate	950/1h	Nanocomposite	5–10	[105]
$Ca_3Co_2O_6$–$Ce_{0.9}Gd_{0.1}O_{2-\delta}$	Sol gel auto-combustion	950–1000	Thin film	500	[136]
$Fe_{0.19}Mn_{0.06}Ce_{0.75}O_{2-\delta}$	Sol gel	1100/4h	Porous faced	36.43	[137]
$La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}$-CuO	Sol gel with EDTA	800	Nanosized crystal	40.67	[138]
$Ca_{3-x}La_xCo_{4-y}Cu_yO_{9+\delta}$-GDC	Solid state reaction	900	Powder	–	[120]
$Ba_{0.5}Sr_{0.5}Co_{0.8}Fe_{0.2}O_3$–$BaZr_{0.65}Ce_{0.2}Y_{0.15}O_3$	Solid reaction method	900	Porous pellet	300–400	[110]
$La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_{3-\delta}$-SDC	Solid state reaction	1350/2h	Pellets	200	[97]
$La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_3-60Gd_{0.1}Ce_{0.9}O_{1.95}$	Glycine-nitrate combustion	975/2h	Powder	200–500	[139]
$La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_{3-\delta}$-NiO	Sol-gel	900	–	179.2	[140]
$Ba_{0.9}La_{0.1}Co_{0.7}Fe_{0.2}Nb0.1O_{3-\delta}$-$Sm_{0.2}Ce_{0.8}O_{2-\delta}$	Electrospinning	950	Fibres	300	[141]
$CoGd_{0.2}Fe_{1.8}O_4$–$Gd_{0.1}Ce_{0.9}O_{2-\delta}$	Sol-gel auto combustion	800/2h	Nanoparticles	75–104	[142]
$La_{0.6}Sr_{0.4}CoO_{3-\delta}$-$Ce_{0.85}Sm_{0.075}Nd_{0.075}O_{2-\delta}$	Solid-liquid method	950/2h	Nano powders	100	[143]
$Bi_{0.5}Sr_{0.5}FeO_{3-\delta}$-$Gd_{0.1}Ce_{0.9}O_{2-\delta}$	Solid state reaction	1100/5h	Coated pellet	400–900	[144]
$Ba_{0.5}Sr_{0.5}Fe_{0.8}Sb_{0.2}O_{3-\delta}$-$Sm_{0.2}Ce_{0.8}O_{2-\delta}$	Co-precipitation-hydrothermal	900/4h	Powder	–	[145]

decrease ohmic loss in SOFC [62,146]. The electrical conductivity must be supported by a considerable amount of ionic (O^{2-}) conductivity, which is accomplished by the vacancy mechanism [147]. In addition to vacancies, geometrical features impact ionic conductivity [148]. As the oxygen vacancy concentration rises and the chemical bonds bend, the electrical conductivity of various composite cathodes diminishes with cation composition range of pure cathode phase (0%) to 50% electrolyte at 700 °C. Increased oxygen vacancy concentration reduces charge carrier concentration and perturbs the periodic potential, resulting in carrier localization [96]. Fig. 3b shows the thermal expansion of different composite materials varying with cation composition (wt.%) at 700 °C. Reduced thermal stress between cell components, such as the electrolyte and connecting materials, can be achieved using a TEC that is compatible with the cell components [149,150]. When the cathode and electrolyte have a mismatched TEC, thermal cycling can create significant strain, leading to cracks and delamination in the components [151]. TEC is determined by the bond strength of the system's elements. TEC is lower in covalently bound compounds like cerium than in metallic and ionic systems [152]. In Fig. 3b, at cation composition between pure phase of cathode and 50% by weight of electrolyte, the slope of the thermal expansion curve abruptly changes. The heat reduction of related ions and the shifting of ions from a low to a high spin state induce these slope changes, typically connected with chemical expansion. This sudden slope shift is lessened when electrolytes are added [96]. As a result, improving thermal expansion fitting between the electrolyte and cathode components may be possible by including an electrolyte in a general cathode [153].

For various composite cathodes, Fig. 3c shows how the electrolyte concentration affects the polarization resistance (R_p) at 700 °C. As shown in Fig. 3c, increased electrolyte content lowers R_p values to a certain wt. % of SDC electrolyte and then rises with the addition of more SDC. The composite cathodes' oxygen ion transportability and TPB zone area are enhanced by the SDC electrolyte's superior oxygen ionic conductivity [154]. As a result, the electrochemical behavior of the general cathode is greatly improved by adding SDC [155]. Fig. 3d, e shows the variations in composite cathode voltage and power density with respect to current density at 700 °C. From the figure, it is clear that introducing the proper amount of electrolyte phase into the available cathode enhances the electrochemical properties of the composite cathodes. This type of composite cathode is better suited for IT-SOFCs than the current base cathodes.

4.2. Chemical compatibility and stability

While SOFC is a clean and efficient electricity generation device, it requires a high working temperature (800–1000 °C). Due to its poor chemical compatibility with cell components and difficult high-temperature sintering process, it has limited commercial applicability [29]. $NdSrCo_2O_{5-\delta}$ (NSCO) is a perovskite with an electrical conductivity of 921.7 S cm^{-1} at 800 °C and has a metal like temperature dependence. To test the chemical compatibility of NSCO ceramic and GDC electrolyte, a 1:1 by wieght mixture was heated to 1100 °C for 2 h. NSCO was post-fired with GDC for SOFC integration, hence its firing temperature determined the heat treatment temperature. The heat treatment composite mixture's Rietveld refined XRD pattern showed no second phase peaks, indicating no chemical reactions within the instrument's detection limit [156]. The phase stability of the cathode's component elements is directly connected to its catalytic activity, working strength, and structural stability under an oxidising environment [70,157]. Powder characteristics like powder size, material composition, and composite cathode morphology (calcination effects, cell design, and cathode structure) significantly impact SOFC performance [158]. The chemical compatibility of the cathodes with the electrolyte is crucial for the long-term durability of SOFCs [159]. Y doping on $SrCoO_{3-\delta}$ (SYC) cathode and LSGM (or GDC) electrolyte chemical compatibility was investigated using the calcination of a mixture of an equal content of SYC10 and LSGM (or GDC) [160,161]. Chemical compatibility with neighboring cell components is necessary for SOFC long-term stability and to prevent unwanted interface reactions [162]. Composite cathodes of BSFO-LSM demonstrated remarkable performance as a type of IT-SOFC cathode. In comparison to pure BSFO cathodes, the BSFO-LSM composite cathode exhibited significantly reduced polarization

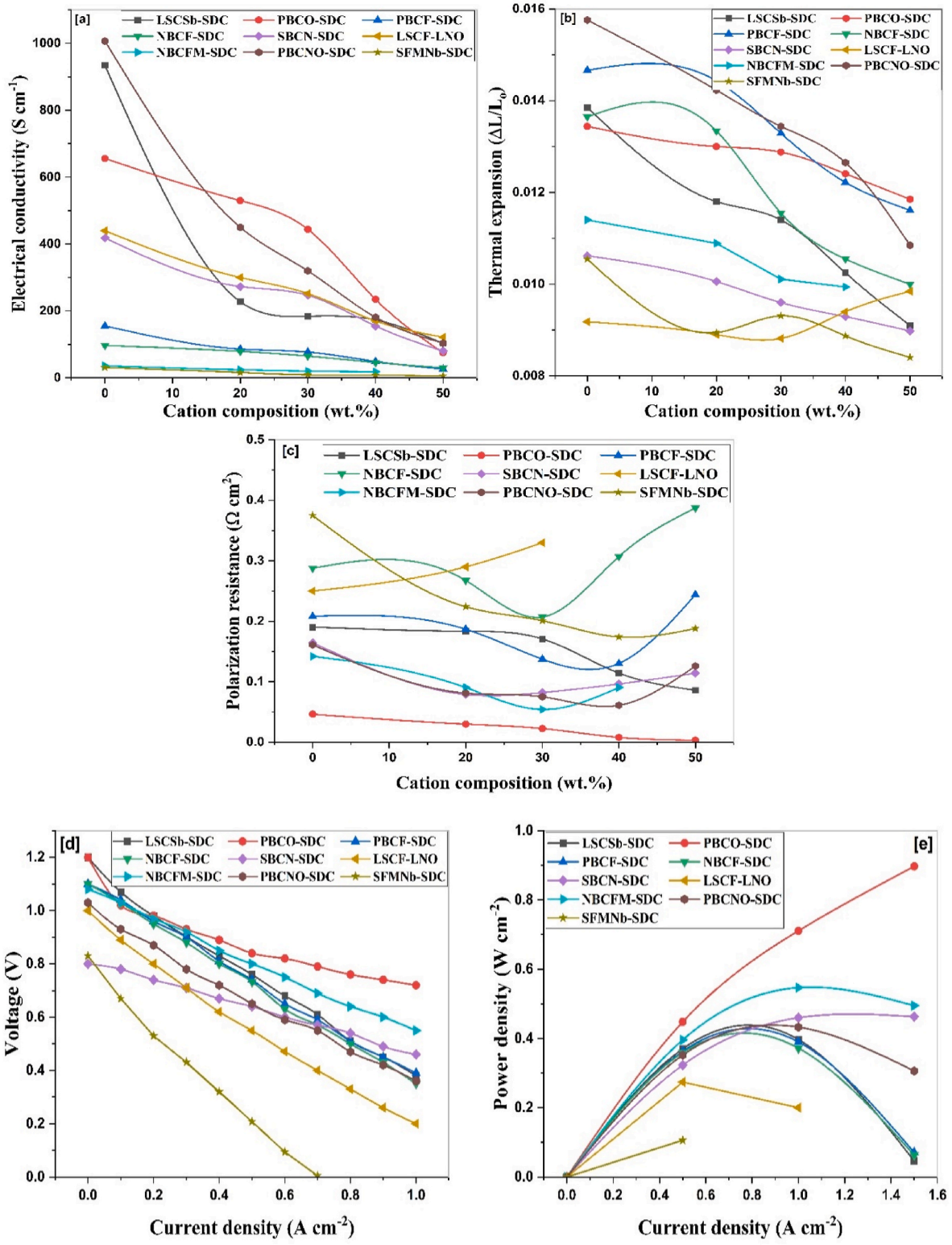


Fig. 3a–e. Analysis of composite cathode properties for selected materials at 700 °C.

resistance, measuring at 0.11 Ω cm^2 at 800 °C. Moreover, the single cell displayed robust stability at 650 °C, maintaining an electric current density of 400 mA cm^{-2} for a continuous duration of 50 h [163].

4.3. Mechanical bonding, porosity level, and sintering activity

Composite cathodes in SOFCs frequently utilize a blend of ionic and electronic conductors to boost overall performance. Mechanical bonding is pivotal in establishing effective contact between various components, ensuring a sturdy and unified structure. A prevalent strategy involves developing a porous structure where particles of diverse materials interlock, enhancing the mechanical integrity of the cathode. This approach contributes to improved stability during thermal cycling and fosters optimal electrical contact between different phases. Key aspects

Table 3
Properties of nine selected composite cathodes at 700 °C.

Single Cell	Electrolyte Thickness (μm)	R_p (Ω cm^2)	MPD (W cm^{-2})	TEC (k^{-1})	Electrical conductivity (S cm^{-1})	Ref.
Ni-YSZ/YSZ/GDC/PBCO-40SDC	4	0.008	0.897	0.01241	234.81	[170]
NiO-SDC/SDC/SBCN-20SDC	700	0.079	0.62	0.0101	273.52	[171]
NiO-SDC/SDC/LSGM/LSCSb-50SDC	300	0.086	0.62	0.0091	103.27	[172]
NiO-SDC/SDC/PBCNO-40SDC	300	0.061	0.48	0.0127	181.56	[173]
NiO-ScSZ/ScSZ/SDC/SFMNb-40SDC	400	0.174	1.17	0.0088	7.7802	[24]
NiO-SDC/YSZ-SDC/LSCF-30LNO	3000	0.33	0.274	0.0088	251.95	[174]
NiO-SDC/SDC/LSGM/PBCF-40SDC	300	0.130	0.95	0.0122	48.79	[175]
NiO-SDC/SDC/LSGM/NBCF-30SDC	300	0.207	0.87	0.0115	64.77	[175]
NiO-YSZ/YSZ/SDC/NBCFM-30SDC	–	0.542	0.547	0.01011	20.525	[176]

of mechanical bonding in composite cathodes encompass particle interlocking, sintering and densification, co-firing, binder materials, microstructure design, and thermal expansion matching [164]. Sintering activity is typically necessary for the cathode-interconnect interface to maintain its mechanical and structural integrity [165]. Jafari et al. [115] investigated the chemical stability of samples using XRD to identify potential reaction products between the LCFN cathode and the YSZ electrolyte in composite phases. To determine an appropriate sintering temperature, pellets of 50LCFN-50YSZ samples were subjected to uniaxial pressure (30 MPa) and sintered in air for nearly 168 h at temperatures of 850, 900, 950, and 1000 °C. They observed the emergence of impurity phases at temperatures exceeding 850 °C, prompting them to limit the sintering temperature to not exceed 850 °C. For the reaction to proceed, the cathode/electrolyte interface must have enough porosity to allow gaseous oxygen to enter the three-phase boundary [65]. The porosity of the cathode impacts gas transport between electrodes, electrolyte ionic conductivity, and, as a result, cell performance [166]. Different components of SOFCs should be able to endure mechanical stresses induced during production and operation. Appropriate mechanical qualities are required to avoid failures such as cracks, delamination, and fractures [167].

4.4. Catalytic activity

The oxygen reduction reaction mechanism has high catalytic activity, which is vital for increasing power output. As a result, electrocatalytic activity is a critical component to consider while developing cathode materials [20,168]. In addition, as opposed to a pure electronic conductor cathode (e.g., Sr-doped $LaMnO_3$), which restricts the ORR to a thin domain at the electrolyte surface, a high MIEC material is ideal for a SOFC cathode, enabling oxygen ion diffusion throughout the entire cathode domain [157]. Li et al. [169] explored the impact of incorporating $Ce_{0.8}Sm_{0.8}O_{1.9}$ with $FeCo_2O_{4-\delta}$ on both physical and electrochemical properties. In contrast to the $FeCo_2O_{4-\delta}$ cathode, the $FeCo_2O_{4-\delta}$-$60Ce_{0.8}Sm_{0.8}O_{1.9}$ composite cathode exhibited enhanced electrocatalytic activity and a higher power density of 397 mW-cm^{-2} at 800 °C.

Based on literature available on composite cathode and their suggestions, weightage to the properties has been given. Like, it has given 30-30% weightage to Rp and MPD and 20-20% weightage to TEC and conductivity. We have made a reasoning of material order based on properties, as depicted in Table 3. SDC electrolytes in the cathode sample prevent the cathode particles from contacting one another. Increased oxygen vacancy concentration reduces charge carrier concentration while also causing the periodic potential to be perturbed. It results in the localization of the carrier. From Table 3, SBCN-20SDC offers maximum electrical conductivity within the cation composition of pure cathode to 50% of electrolyte at 700 °C, while SFMNb-40SDC gives the lowest electrical conductivity. The expansion test reveals that when SDC is added to cathode material, the TEC is reduced, and the TPB area is expanded, which improves thermal expansion matching between the cathode and electrolyte. With increasing cation composition from 0 to 50% at 700 °C, the difference between composite electrode and electrolyte is minimum for LSCF-30SDC and SFMNb-40SDC and followed by LSCSb-50SDC, SBCN-20SDC, NBCFM-30SDC, NBCF-30SDC, PBCF-40SDC, PBCO-40SDC, and PBCNO-40SDC respectively, which is due to the thermal reduction of associated ions and the transition of ions from the low to high spin rate.

Polarization resistance values decrease with increasing content of SDC up to a certain wt% of electrolyte and then increase upon further addition of electrolyte. Among these composite cathodes, PBCO-40SDC has the lowest R_p value, which enhances the cathode's oxygen ionic conductivity and expands the TPB zone in contact with the PBCO cathode. The cathode's capacity to conduct oxygen ions with the introduction of electrolyte content in descending order is PBCO-40SDC, PBCNO-40SDC, SBCN-20SDC, LSCSb-50SDC, PBCF-40SDC, SFMNb-40SDC, NBCF-30SDC, LSCF-30LNO, and NBCFM-30SDC respectively. The maximum power density of the cells is offered by SFMNb-40SDC followed by PBCF-40SDC, PBCO-40SDC, NBCF-30SDC, LSCSb-50SDC, SBCN-20SDC, NBCFM-30SDC, PBCNO-40SDC, and LSCF-30LNO respectively, at 700 °C. It is due to the enhanced ORR activity and extended TPB sites by SFMNb-40SDC compared to other composite cathode materials. Besides the above-discussed composite cathode materials, some other are discussed in Table 4. Investigation of all electrochemical properties like conductivity, TEC, Rp, Voltage, and MPD of these materials has not been covered previously.

5. Conclusions

Solid oxide fuel cells have gained interest as a next-generation power generation system due to their high efficiency (>50%), fuel diversity, and low emissions. It is critical to attain a maximum power density and long-term stability at intermediate temperatures (600–800 °C) for the practical application of SOFCs. The commercialization of SOFCs has been hindered by inadequate catalytic activity and significant degradation of the cathode at intermediate temperatures. Existing SOFCs, on the other hand, have issues including interfacial diffusion, polarization resistance, thermal mismatch, and substantial electrode overpotential. The characteristics of the material, synthesis techniques, and operating conditions all have a significant impact on the kinetics of the intermediate reactions that reduce oxygen. The presence of mixed ionic and electronic conduction (MIEC) within the composite cathode has the potential to modify oxygen stoichiometry and various other physiochemical properties. MIEC is promoting the electrocatalytic functionality of ORR at low operational temperatures (<700 °C). Cobalt-based materials have strong catalytic activity but a low thermal expansion coefficient and chemical stability compared to Fe-based cathodes. Fe-based perovskite, such as $Ln_{1-x}Sr_xFeO_3$ (Ln = La, Pr, Nd, Sm, and Gd), has been studied as IT-SOFC electrode materials due to its good catalytic activity for the oxygen reduction reaction, excellent chemical stability, adequate TEC with the electrolyte, and low cost compared to cobalt-containing materials. Numerous composite cathode materials have been suggested and examined by incorporating electrolyte particles or metal catalysts into the oxide cathode material, as well as altering the cathode microstructure using various cathode techniques. The prevailing approach for fabricating a composite cathode involves the physical

Table 4
Review of other composite cathodes for IT-SOFC at 700 °C.

Composite material	Phase mixing ratio	Polarization resistance (Ω cm^2)	Maximum power density (mW cm^{-2})	Variation compared with pure phase	Ref.
$LaSrCoO_4$–$Ce_{0.9}Mn_{0.1}O_2$	65:35	0.31	–	Rp is reduced by 81.8%. Conductivity is reduced by 48.14%	[55]
$NdSrCO_2O_{5+\delta}$-$Ce_{0.8}Gd_{0.2}O_{2-\delta}$	50:50	0.09	460	TEC is reduced by 47.88%	[156]
$LaNi_{0.6}Fe_{0.4}O_{3-\delta}$-$Bi_{1.5}Y_{0.5}O_3$	50:50	0.53	650	–	[177]
$YBa_2Cu_3O_7$-δ-$La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_{3-\delta}$	70:30	0.044	–	Rp is reduced by 67.88%	[178]
$NdBaCo_2O_{5+\delta}$-$Ce_{0.9}Gd_{0.1}O_{1.95}$	95:5	0.034	1000	Conductivity is reduced by 45.03%. TEC is reduced by 2.87%	[87]
$Ce_{0.9}Pr_{0.1}O_{2-\delta}$-$La_{0.4}Pr_{0.1}Ba_{0.5}Co_{0.975}O_{3-\delta}$	70:30	0.03	–	TEC is reduced by 30.07%	[179]
$La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_{3-\delta}$-$Gd_{0.2}Ce_{0.8}O_2$	60:40	–	178	–	[180]
$CuFe_2O_4$–$Gd_{0.1}Ce_{0.9}O_{1.95}$	40:60	1.17 (800 °C)	384 (800 °C)	–	[181]
$La_{0.65}Ca_{0.35}FeO_{3-\delta}$-$Gd_{0.2}Ce_{0.8}O_{2-\delta}$	60:40	0.28 (750 °C)	1172 (750 °C)	Rp is reduced by 22.2%	[182]
$La_{0.6}Sr_{0.4}CoO_{3-\delta}$-$Ce_{0.85}Sm_{0.075}Nd_{0.075}O_{2-\delta}$	50:50	0.0597	1013	Rp is reduced by 25%. TEC is reduced by 36.59%	[143]
$CoGd_{0.8}Fe_{1.80}O_4$–$Gd_{0.1}Ce_{0.9}O_{2-\delta}$	–	0.60 (550 °C)	611 (550 °C)	MPD is improved by 28.63%	[142]
$La_{0.6}Sr_{0.4}CoO_{3-\delta}$-$Gd_{0.1}Ce_{0.9}O_{1.95}$	50:50	0.16	796	MPD is improved by 20–35%	[183]
$Ba_{0.5}Sr_{0.5}Fe_{0.8}Sb_{0.2}O_{3-\delta}$-$Sm_{0.2}Ce_{0.8}O_{2-\delta}$	80:20	0.64	890	–	[155]
$Pr_2NiO_{4+\delta}$-$La_{0.6}Sr_{0.4}CoO_{3-\delta}$	–	0.024	618	Rp is reduced by 83.4%	[103]
$Pr_{0.5}Ba_{0.5}Co_{0.7}Fe_{0.3}O_{3+\delta}$-BZCY	–	0.0396	1304	–	[184]
Co_3O_4-GDC-ScCeSZ	–	0.19 (800 °C)	1382 (800 °C)	Rp is reduced by 56%	[105]
$La_2NiO_{4+\delta}$-$Ce_{0.55}La_{0.45}O_{2-\delta}$	70:30	0.21	53	–	[185]
$Pr_{1.95}Ce_{0.05}CuO_4$–$Ce_{0.8}Pr_{0.2}O_{1.9}$	70:30	0.04	1040	Rp is reduced by 71.42%. TEC is reduced by 31.3%	[186]
$La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_{3-\delta}$-$Gd_{0.2}Ce_{0.8}O_{2-\delta}$	50:50	0.45 (650 °C)	600 (650 °C)	Rp is reduced by 77.5%	[187]
$La_{0.84}Sr_{0.16}MnO_{3-\delta}$-$Ce_{0.8}Gd_{0.2}O_{2-\delta}$	70:30	0.149	860	–	[69]
LSCF-BZCY-NBSCF	–	0.14	696	Rp is reduced by 68.18%. MPD is improved by 43.2%	[188]
$La_{0.7}Sr_{0.3}FeO_{3-\delta}$-$Sm_{0.2}Ce_{0.8}O_{2-\delta}$	70:30	0.066	1180	–	[189]
$Ba_{0.5}Sr_{0.5}Fe_{1-x}Cu_xO_{3-\delta}$-$Ce_{0.8}Sm_{0.15}Ca_{0.05}O_{1.875}$	60:40	0.338	–	–	[190]
$La_{0.7}Sr_{0.3}FeO_{3-\delta}$-$BaZr_{0.1}Ce_{0.7}Y_{0.2}O_{3-\delta}$	70:30	0.12	849	–	[191]
$LaMnO_3$-GDC-$CuBi_2O_4$	80:20	0.11	1460	Rp is reduced by 95.21%	[192]
$CuMn_{1.5}Ni_{0.5}O_4$–$Gd_{0.1}Ce_{0.9}O_{2-\delta}$	50:50	0.123 (750 °C)	1076.3 (750 °C)	TEC is reduced by 2.59%	[39]
$LaNi_{0.6}Fe_{0.4}O_{3-\delta}$-$Sm_{0.5}Sr_{0.5}CoO_{3-\delta}$	70:30	0.10	636	–	[193]
$Bi_{0.5}Sr_{0.5}FeO_{3-\delta}$-$Ce_{0.9}Gd_{0.1}O_{1.95}$	70:30	0.14	709	MPD is improved by 71.42%	[194]
$SrCo_{0.8}Nb_{0.1}Ta_{0.1}O_{3-\delta}$-$La_2NiO_4$	90:10	0.12	279	Rp is reduced by 33%.	[40]
$Sm_{0.5}Sr_{0.5}CoO_{3-\delta}$-$Co_3O_4$	70:30	0.069	1037	MPD is improved by 58.33%	[195]

blending of two or more distinct materials. This method is relatively straightforward to execute, although it necessitates strong chemical compatibility among the various constituent materials. The progress in research on composite cathode material for SOFC involves the comprehension of degradation methods, coating and infiltration techniques, doping approaches, nanostructuring, and microstructure control. The absence of in-situ characterization during the operation of SOFCs impedes a more profound comprehension of the contribution of composite cathode technologies to the enhancement mechanisms of SOFCs. A significant and recent advancement is the exsolution of catalyst nanoparticles, which is achieved by using a single precursor phase rather than combining different phases. Polarization resistance, maximum power density, thermal expansion coefficient, and electrical conductivity of nine selected materials PBCO-40SDC, PBCNO-40SDC, SBCN-20SDC, LSCSb-50SDC, PBCF-40SDC, SFMNb-40SDC, NBCF-30SDC, LSCF-30LNO, and NBCFM-30SDC has been done. Outcomes show that The cathode's capacity to conduct oxygen ions with the introduction of electrolyte content in descending order is PBCO-40SDC, PBCNO-40SDC, SBCN-20SDC, LSCSb-50SDC, PBCF-40SDC, SFMNb-40SDC, NBCF-30SDC, LSCF-30LNO, and NBCFM-30SDC respectively. The PBCO-40SDC cathode has the lowest polarization resistance compared to other composite cathodes. It also has a high maximum power density, maximum electrical conductivity, and low TEC.

6. Future research recommendations

- Future research should focus on infiltrated electrode series resistance, especially those with extremely porous backbones that only conduct ions.
- Exsolution of catalytically active nanoparticles (from a single precursor phase, instead of mixing separate phases) is a recent and important development.
- There are still gaps in the improvement of material research and development, various techniques for fabrication of cathode materials, cathode contact layer performance, etc.
- In the future, it is crucial to prioritise the examination of the cost and enduring stability of composite cathode materials.

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.

Acknowledgments

Anil Kumar Yadav expresses gratitude to Dr. A.P.J. Abdul Kalam Technical University, Lucknow, Uttar Pradesh, India, for awarding the fellowship as part of the Homi Bhabha Teaching cum Research Fellowship. The authors extend their appreciation for the support received from the Department of Mechanical Engineering at the Institute of Engineering and Technology, Lucknow, India, as well as the encouragement provided by the Centre for Energy and Environment at Delhi Technological University, Delhi, India.

References

[1] Zhang L, Jia C, Bai F, Wang W, An S, Zhao K, et al. A comprehensive review of the promising clean energy carrier: hydrogen production, transportation, storage,

and utilization (HPTSU) technologies. Fuel 2024;355:129455. https://doi.org/10.1016/j.fuel.2023.129455.
[2] Kumar R, Kumar A, Pal A. Overview of hydrogen production from biogas reforming: technological advancement. Int J Hydrogen Energy 2022;47:34831–55. https://doi.org/10.1016/j.ijhydene.2022.08.059.
[3] Stambouli AB, Traversa E. Solid oxide fuel cells (SOFCs): a review of an environmentally clean and efficient source of energy. Renew Sustain Energy Rev 2002;6:433–55. https://doi.org/10.1016/S1364-0321(02)00014-X.
[4] Shafiee S, Topal E. When will fossil fuel reserves be diminished? Energy Pol 2009;37:181–9. https://doi.org/10.1016/j.enpol.2008.08.016.
[5] Sreedhar I, Agarwal B, Goyal P, Singh SA. Recent advances in material and performance aspects of solid oxide fuel cells. J Electroanal Chem 2019;848:113315. https://doi.org/10.1016/j.jelechem.2019.113315.
[6] Choudhary T, Sahu MK. Renewable energy and its innovative technologies. Springer Singapore; 2019. https://doi.org/10.1007/978-981-13-2116-0.
[7] Muhammed NS, Gbadamosi AO, Epelle EI, Abdulrasheed AA, Haq B, Patil S, et al. Hydrogen production, transportation, utilization, and storage: recent advances towards sustainable energy. J Energy Storage 2023;73:109207. https://doi.org/10.1016/j.est.2023.109207.
[8] Kumar Yadav A, Sinha S, Kumar A. Comprehensive review on performance assessment of solid oxide fuel cell-based hybrid power generation system. Therm Sci Eng Prog 2023;46:102226. https://doi.org/10.1016/j.tsep.2023.102226.
[9] Ahmed S, Krumpelt M. Hydrogen from hydrocarbon fuels for fuel cells. Int J Hydrogen Energy 2001;26:291–301. https://doi.org/10.1016/S0360-3199(00)00097-5.
[10] Park S, Vohs JM, Gorte RJ. Direct oxidation of hydrocarbons in a solid-oxide fuel cell. Nature 2000;404:265–7. https://doi.org/10.1038/35005040.
[11] Xu Q, Xia L, He Q, Guo Z, Ni M. Thermo-electrochemical modelling of high temperature methanol-fuelled solid oxide fuel cells. Appl Energy 2021;291:116832. https://doi.org/10.1016/j.apenergy.2021.116832.
[12] Zhang M, Jeerh G, Zou P, Lan R, Wang M, Wang H, et al. Recent development of perovskite oxide-based electrocatalysts and their applications in low to intermediate temperature electrochemical devices. Mater Today 2021;49:351–77. https://doi.org/10.1016/j.mattod.2021.05.004.
[13] P K, Bm S, Bn W, Pvb R. Review on different components of solid oxide fuel cells. J Powder Metall Min 2017;6. https://doi.org/10.4172/2168-9806.1000181.
[14] Abdalla AM, Hossain S, Petra PMI, Ghasemi M, Azad AK. Achievements and trends of solid oxide fuel cells in clean energy field: a perspective review. Front Energy 2020;14:359–82. https://doi.org/10.1007/s11708-018-0546-2.
[15] Yilmaz F, Ozturk M. Design and modeling of an integrated combined plant with SOFC for hydrogen and ammonia generation. Int J Hydrogen Energy 2022;47:31911–26. https://doi.org/10.1016/j.ijhydene.2022.01.249.
[16] Vostakola MF, Horri BA. Progress in material development for low-temperature solid oxide fuel cells: a review. Energies 2021;14. https://doi.org/10.3390/en14051280.
[17] Alzahrani A, Dincer I, Li X. A performance assessment study on solid oxide fuel cells for reduced operating temperatures. Int J Hydrogen Energy 2015;40:7791–7. https://doi.org/10.1016/j.ijhydene.2014.11.115.
[18] Hussain S, Yangping L. Review of solid oxide fuel cell materials: cathode, anode, and electrolyte. Energy Transitions 2020;4:113–26. https://doi.org/10.1007/s41825-020-00029-8.
[19] Bello IT, Zhai S, He Q, Xu Q, Ni M. Scientometric review of advancements in the development of high-performance cathode for low and intermediate temperature solid oxide fuel cells: three decades in retrospect. Int J Hydrogen Energy 2021;46:26518–36. https://doi.org/10.1016/j.ijhydene.2021.05.134.
[20] Vinoth Kumar R, Khandale AP. A review on recent progress and selection of cobalt-based cathode materials for low temperature-solid oxide fuel cells. Renew Sustain Energy Rev 2022;156:111985. https://doi.org/10.1016/j.rser.2021.111985.
[21] Gu H, Xu M, Song Y, Zhou C, Su C, Wang W, et al. SrCo0.8Ti0.1Ta0.1O3-δ perovskite: a new highly active and durable cathode material for intermediate-temperature solid oxide fuel cells. Composites Part B 2021;213:108726. https://doi.org/10.1016/j.compositesb.2021.108726.
[22] Hossain S, Abdalla AM, Jamain SNB, Zaini JH, Azad AK. A review on proton conducting electrolytes for clean energy and intermediate temperature-solid oxide fuel cells. Renew Sustain Energy Rev 2017;79:750–64. https://doi.org/10.1016/j.rser.2017.05.147.
[23] Abd Aziz AJ, Baharuddin NA, Somalu MR, Muchtar A. Review of composite cathodes for intermediate-temperature solid oxide fuel cell applications. Ceram Int 2020;46:23314–25. https://doi.org/10.1016/j.ceramint.2020.06.176.
[24] Guan M, Sun W, Ren R, Fan Q, Qiao J, Wang Z, et al. Improved electrochemical performance of Sr2Fe1.5Mo0.4Nb0.1O6-δ -Sm0.2Ce0.8O2-δ composite cathodes by a one-pot method for intermediate temperature solid oxide fuel cells. Int J Hydrogen Energy 2016;41:3052–61. https://doi.org/10.1016/j.ijhydene.2015.12.056.
[25] Rembelski D, Viricelle JP, Combemale L, Rieu M. Characterization and comparison of different cathode materials for SC-SOFC: LSM, BSCF, SSC, and LSCF. Fuel Cell 2012;12:256–64. https://doi.org/10.1002/fuce.201100064.
[26] da Silva FS, de Souza TM. Novel materials for solid oxide fuel cell technologies: a literature review. Int J Hydrogen Energy 2017;42:26020–36. https://doi.org/10.1016/j.ijhydene.2017.08.105.
[27] Tarancón A, Burriel M, Santiso J, Skinner SJ, Kilner JA. Advances in layered oxide cathodes for intermediate temperature solid oxide fuel cells. J Mater Chem 2010;20:3799–813. https://doi.org/10.1039/b922430k.
[28] Burnwal SK, Bharadwaj S, Kistaiah P. Review on MIEC cathode materials for solid oxide fuel cells. J Mol Eng Mater 2016;4:1630001. https://doi.org/10.1142/s2251237316300011.
[29] Iqbal M, Muneer M, Raza R, Jamal MA. Preparation of composite cathode material by using the extracted lead (Pb) of waste lead acid battery for LT-SOFC. Ceram Int 2022;48:19681. https://doi.org/10.1016/j.ceramint.2022.03.105. –7.
[30] Subhashini P VCK, Rajesh K VD. Doped BSCF cathode materials for low-temperature solid oxide fuel cell applications-A short review. Mater Today Proc 2023;78:520–3. https://doi.org/10.1016/j.matpr.2022.11.212.
[31] Chen F, Zhou D, Xiong X, Pan J, Cai D, Wei Z, et al. Doping strategy on improving the overall cathodic performance of double perovskite LnBaCo2O5+δ (Ln=Pr, Gd) as potential SOFC cathode materials. J Mater 2023. https://doi.org/10.1016/j.jmat.2023.02.004.
[32] Qi H, Zhang T, Qiu P, Liu D, Tu B, Cheng M. Comparison of nanostructured composite cathodes synthesized by liquid self-assembly and nanosolid mechanical-mixing for solid oxide fuel cell. Int J Hydrogen Energy 2022;47:15721–30. https://doi.org/10.1016/j.ijhydene.2022.03.053.
[33] Junaida A, Aziz A, Baharuddin NA, Somalu MR, Muchtar A. Layering optimization of the SrFe0.9 Ti0.1O3−δ –Ce0.8Sm0.2O1.9 composite cathode. Molecules 2022:1–11. https://doi.org/10.3390/molecules27082549.
[34] Juhl M, Primdahl S, Manon C, Mogensen M. Performance/structure correlation for composite SOFC cathodes. J Power Sources 1996;61:173–81. https://doi.org/10.1016/S0378-7753(96)02361-0.
[35] Zeng D, Xu K, Zhu F, Chen Y. Enhancing the oxygen reduction reaction activity and durability of a solid oxide fuel cell cathode by surface modification of a hybrid coating. Int J Hydrogen Energy 2023;48:23992–4001. https://doi.org/10.1016/j.ijhydene.2023.03.198.
[36] Yurchenko MV, Antonova EP, Tropin ES, Suntsov AY. Adjusting electrochemical properties of PrBaCo2O6–δ as SOFC cathode by controllable Ca3Co4O9 additions. Ceram Int 2023;49:21485–91. https://doi.org/10.1016/j.ceramint.2023.03.281.
[37] Song KE, Schlegl H, Kang H, Choi W, Kim JH. Electrochemical characteristic of non-stoichiometric SmBa0.45Sr0.5Co2O5+d layered perovskite oxide system for IT-SOFC cathode. Int J Hydrogen Energy 2023;48:17664–76. https://doi.org/10.1016/j.ijhydene.2023.01.255.
[38] Xia B, Zhang H, Yao C, Lou H, Chen M, Zhang Z, et al. Enhancing ORR activity and CO2 tolerance of Pr0.4Sr0.6Co0.2Fe0.8O3-δ-based SOFC cathode through synergistic doping and surface modification. Appl Surf Sci 2024;649:159143. https://doi.org/10.1016/j.apsusc.2023.159143.
[39] Duan Y, Cao B, Li S, Wu A, Huang Z, Wang J, et al. Properties of CuMn1.5Ni0.5O4 spinel as high-performance cathode for solid oxide fuel cells. J Eur Ceram Soc 2023;43:5298–306. https://doi.org/10.1016/j.jeurceramsoc.2023.05.003.
[40] Li J, Gao Z, Min H, Li M, Lu Y, Wang X, et al. Enabled fast cathode kinetics for intermediate-temperature solid oxide fuel cell with improved CO2 poisoning robustness: La2NiO4 surfaced-modified SrCo0.8Nb0.1Ta0.1O3-δ composite. J Power Sources 2021;506. https://doi.org/10.1016/j.jpowsour.2021.230057.
[41] Xu X, Cao C, Xia C, Peng D. Electrochemical performance of LSM-SDC electrodes prepared with ion-impregnated LSM. Ceram Int 2009;35:2213–8. https://doi.org/10.1016/j.ceramint.2008.12.001.
[42] Hu L, Zhou D, Zhu X, Wang N, Bai J, Gong H, et al. A high-performance composite cathode based on thermal expansion complementation for SOFC. Fuel 2024;362:130864. https://doi.org/10.1016/j.fuel.2024.130864.
[43] Adler SB, Lane JA, Steele BCH. Electrode kinetics of porous mixed-conducting oxygen electrodes. J Electrochem Soc 1996;143:3554–64. https://doi.org/10.1149/1.1837252.
[44] Ndubuisi A, Abouali S, Singh K, Thangadurai V. Recent advances, practical challenges, and perspectives of intermediate temperature solid oxide fuel cell cathodes. J Mater Chem A 2022;10:2196–227. https://doi.org/10.1039/d1ta08475e.
[45] Geofrey Sahini M, Daud Lupyana S. Perspective and control of cation interdiffusion and interface reactions in solid oxide fuel cells (SOFCs). Mater Sci Eng B 2023;292:116415. https://doi.org/10.1016/j.mseb.2023.116415.
[46] Li Y, Gemmen R, Liu X. Oxygen reduction and transportation mechanisms in solid oxide fuel cell cathodes. J Power Sources 2010;195:3345–58. https://doi.org/10.1016/j.jpowsour.2009.12.062.
[47] Sun C, Hui R, Roller J. Cathode materials for solid oxide fuel cells: a review. J Solid State Electrochem 2010;14:1125–44. https://doi.org/10.1007/s10008-009-0932-0.
[48] Irshad M, Siraj K, Raza R, Ali AA, Tiwari P, Zhu B, et al. A brief description of high temperature solid oxide fuel cell's operation, materials, design, fabrication technologies and performance. Appl Sci 2016;6. https://doi.org/10.3390/app6030075.
[49] Park BK, Lee SB, Lim TH, Song RH, Lee JW. High-performance solid oxide fuel cell with an electrochemically surface-tailored oxygen electrode. ChemSusChem 2018;11:2620–7. https://doi.org/10.1002/cssc.201800962.
[50] Zhou W, Ran R, Shao Z. Progress in understanding and development of Ba0.5Sr0.5Co0.8Fe0.2O3-δ-based cathodes for intermediate-temperature solid-oxide fuel cells: a review. J Power Sources 2009;192:231–46. https://doi.org/10.1016/j.jpowsour.2009.02.069.
[51] Kim D, Lee KT. Effect of lanthanide (Ln=La, Nd, and Pr) doping on electrochemical performance of Ln2NiO4+δ−YSZ composite cathodes for solid oxide fuel cells. Ceram Int 2021;47:2493–8. https://doi.org/10.1016/j.ceramint.2020.09.092.
[52] Lesani P, Babaei A, Ataie A, Mostafavi E. Nanostructured MnCo2O4 synthesized via co-precipitation method for SOFC interconnect application. Int J Hydrogen Energy 2016;41:20640–9. https://doi.org/10.1016/j.ijhydene.2016.07.216.

[53] Ahmad MZ, Ahmad SH, Chen RS, Ismail AF, Hazan R, Baharuddin NA. Review on recent advancement in cathode material for lower and intermediate temperature solid oxide fuel cells application. Int J Hydrogen Energy 2022;47:1103–20. https://doi.org/10.1016/j.ijhydene.2021.10.094.

[54] Zheng Y, Zhao C, Wu T, Li Y, Zhang W, Zhu J, et al. Enhanced oxygen reduction kinetics by a porous heterostructured cathode for intermediate temperature solid oxide fuel cells. Energy AI 2020;2:100027. https://doi.org/10.1016/j.egyai.2020.100027.

[55] Wu K, Fu L, Zhou J, Zhou L, Yang J, Liu Z, et al. Enhanced oxygen reduction reaction of LaSrCoO4–Ce0.9Mn0.1O2 composite cathode for solid oxide fuel cells. Ceram Int 2023;49:9848–56. https://doi.org/10.1016/j.ceramint.2022.11.158.

[56] Hjalmarsson P, Mogensen M. La0.99Co0.4Ni0.6O3-δ- Ce0.8Gd0.2O1.95 as composite cathode for solid oxide fuel cells. J Power Sources 2011;196:7237–44. https://doi.org/10.1016/j.jpowsour.2010.11.065.

[57] Yasin L, Atkinson A, Cooper SJ, Bertei A. Identifiability of the mechanisms governing the reaction kinetics of MIEC electrodes in solid oxide cells. Electrochim Acta 2023;472:143418. https://doi.org/10.1016/j.electacta.2023.143418.

[58] Pichandi R, Tiwari P, Basu S, Singh S. Ca2Fe2O5 and Ca2Co2O5 based composite cathode for IT-SOFC application. Solid State Ionics 2024;404:116418. https://doi.org/10.1016/j.ssi.2023.116418.

[59] Desta HG, Tian D, Yang Q, Zhu S, Song K, Chen Y, et al. Developing a new Sr and Co-free composite cathode of solid oxide fuel cells with high performance. Chem Phys Lett 2022;806:140037. https://doi.org/10.1016/j.cplett.2022.140037.

[60] Yu J, Cheng L, Luo L, Wang L, Xu X, Liu S, et al. Surface modulated B-site doping of PrBa0.5Sr0.5Co2-xFexO5+δ as highly efficient cathode for intermediate-temperature solid oxide fuel cells. Solid State Ionics 2023:402. https://doi.org/10.1016/j.ssi.2023.116380.

[61] Liu Y, Kang K, Pan Z, Wang C, Jiang K, Wang Y. Enhancing the oxygen reduction reaction activity of SOFC cathode via construct a cubic fluorite/perovskite heterostructure. Appl Surf Sci 2024;642:158405. https://doi.org/10.1016/j.apsusc.2023.158405.

[62] Gao Y, Yang Y, Lin X, Fu M, Hu W, Tong H, et al. Investigation and study of three different composite cathodes for proton-conducting solid oxide fuel cells. Sep Purif Technol 2022;300:121890. https://doi.org/10.1016/j.seppur.2022.121890.

[63] Balaguer M, Vert VB, Navarrete L, Serra JM. SOFC composite cathodes based on LSM and co-doped cerias (Ce 0.8Gd 0.1X 0.1O 2-δ, X = Gd, Cr, Mg, Bi, Ce). J Power Sources 2013;223:214–20. https://doi.org/10.1016/j.jpowsour.2012.09.060.

[64] Yang Y, Liu X, Wang P, Teng Y, Chu X, Li J, et al. Effective calcium co-doping in Sm0.8Ca0.2Ba1-xCaxCo2O5+δ cathode materials for intermediate-temperature solid oxide fuel cells. Ceram Int 2023;49:40358–65. https://doi.org/10.1016/j.ceramint.2023.10.008.

[65] Wang M, Su C, Zhu Z, Wang H, Ge L. Composite cathodes for protonic ceramic fuel cells: rationales and materials. Composites Part B 2022;238:109881. https://doi.org/10.1016/j.compositesb.2022.109881.

[66] Ouyang Y, Zhu D, Zhu C, Zhang Y, Liu J, Jia X, et al. Composite electrolyte with Ruddlesden-Popper structure Sm1.2Sr0.8Ni0.6Fe0.4O4+δ for high-performance low temperature solid oxide fuel cells. Int J Hydrogen Energy 2023;48:268–79. https://doi.org/10.1016/j.ijhydene.2022.09.204.

[67] Gao Y, Zhang M, Fu M, Hu W, Tong H, Tao Z. A comprehensive review of recent progresses in cathode materials for Proton-conducting SOFCs. Energy Rev 2023;2:100038. https://doi.org/10.1016/j.enrev.2023.100038.

[68] Gür TM. Comprehensive review of methane conversion in solid oxide fuel cells: prospects for efficient electricity generation from natural gas. Prog Energy Combust Sci 2016;54:1–64. https://doi.org/10.1016/j.pecs.2015.10.004.

[69] Li J, Shi J, Nie H, Zhan Z, Wang S. Improvement of a GDC-based composite cathode for intermediate-temperature solid oxide fuel cells. J Electroceram 2014;32:339–43. https://doi.org/10.1007/s10832-014-9909-1.

[70] Lee S, Miller N, Gerdes K. Long-term stability of SOFC composite cathode activated by electrocatalyst infiltration. J Electrochem Soc 2012;159:F301–8. https://doi.org/10.1149/2.067207jes.

[71] Jiang Z, Xia C, Chen F. Nano-structured composite cathodes for intermediate-temperature solid oxide fuel cells via an infiltration/impregnation technique. Electrochim Acta 2010;55:3595–605. https://doi.org/10.1016/j.electacta.2010.02.019.

[72] Fan L, Wang Y, Jia Z, Xiong Y, Brito ME. Nanofiber-structured SSC-GDC composite cathodes for a LSGM electrolyte based IT-SOFCs. Ceram Int 2015;41:6583–8. https://doi.org/10.1016/j.ceramint.2015.01.104.

[73] Cheng J, Qian W, Wang P, Tian C. A high activity cathode of Sm0.2Ce0.8O1.9 decorated Mn1.5Co1.5O4 using ion impregnation technique within a solid oxide fuel cell system. Solid State Sci 2022;131:106962. https://doi.org/10.1016/j.solidstatesciences.2022.106962.

[74] Long Q, Sha R, Wang R, Xu B, Men H, Wang Q, et al. Research progress of composite cathode materials for Solid oxide fuel cells. Prog Nat Sci Mater Int 2023;33:267–78. https://doi.org/10.1016/j.pnsc.2023.08.016.

[75] Gao J, Meng X, Luo T, Wu H, Zhan Z. Symmetrical solid oxide fuel cells fabricated by phase inversion tape casting with impregnated SrFe0.75Mo0.25O3-Δ (SFMO) electrodes. Int J Hydrogen Energy 2017;42:18499–503. https://doi.org/10.1016/j.ijhydene.2017.03.205.

[76] Liu Y, Pan Z, Zhao X, Zhong S, Chen X, Wang C. Performance of PrBaCo2O5+δ impregnated La0.6Sr0.4Co0.2Fe0.8O3−δ composites as potential cathode materials for solid oxide fuel cells. J Alloys Compd 2021;886:161155. https://doi.org/10.1016/j.jallcom.2021.161155.

[77] Jiang SP, Wang W. Fabrication and performance of GDC-impregnated (La,Sr)MnO [sub 3] cathodes for intermediate temperature solid oxide fuel cells. J Electrochem Soc 2005;152:A1398. https://doi.org/10.1149/1.1928167.

[78] Zarabi Golkhatmi S, Asghar MI, Lund PD. A review on solid oxide fuel cell durability: latest progress, mechanisms, and study tools. Renew Sustain Energy Rev 2022;161:112339. https://doi.org/10.1016/j.rser.2022.112339.

[79] Singh M, Zappa D, Comini E. Solid oxide fuel cell: decade of progress, future perspectives and challenges. Int J Hydrogen Energy 2021;46:27643–74. https://doi.org/10.1016/j.ijhydene.2021.06.020.

[80] Meng X, Wang Y, Zhao Y, Zhang T, Yu N, Chen X, et al. In-situ exsolution of nanoparticles from Ni substituted Sr2Fe1.5Mo0.5O6 perovskite oxides with different Ni doping contents. Electrochim Acta 2020;348:1–12. https://doi.org/10.1016/j.electacta.2020.136351.

[81] Pang S, Yang G, Xu J, Zhu M, Wang W, Shen X. Characterization of SmBa0.5Ca0.5CoCuO5+δ with exsolved CaCuO2-type phase as an advanced cathode material of solid oxide fuel cells. Ceram Int 2020;46:3861–5. https://doi.org/10.1016/j.ceramint.2019.10.111.

[82] Yang Z, Zhang J, Wu B, Zhao X, Lu X, Zhao Y, et al. Self-assembled La0.6Sr0.4FeO3-δ-La1.2Sr0.8NiO4+δ composite cathode for protonic ceramic fuel cells. Ceram Int 2023;49:25381–8. https://doi.org/10.1016/j.ceramint.2023.05.074.

[83] Qiu P, Yang X, Zhu T, Sun S, Jia L, Li J. Review on core-shell structured cathode for intermediate temperature solid oxide fuel cells. Int J Hydrogen Energy 2020;45:23160–73. https://doi.org/10.1016/j.ijhydene.2020.06.034.

[84] Yang D, Chen G, Liu H, Zhang L, He Y, Zhang X, et al. Electrochemical performance of a Ni0.8Co0.15Al0.05LiO2 cathode for a low temperature solid oxide fuel cell. Int J Hydrogen Energy 2021;46:10438–47. https://doi.org/10.1016/j.ijhydene.2020.06.189.

[85] Mushtaq N, Xia C, Dong W, Wang B, Raza R, Ali A, et al. Tuning the energy band structure at interfaces of the SrFe0.75Ti0.25O3-δ-Sm0.25Ce0.75O2-δ heterostructure for fast ionic transport. ACS Appl Mater Interfaces 2019;11:38737–45. https://doi.org/10.1021/acsami.9b13044.

[86] Qi H, Zhang T, Liu D, Cheng M, Tu B. Investigation of a self-assembled Sr0.74La0.26CoO3-δ-SrZr0.79Co0.21O3-δ composite with hierarchical structure as intermediate-temperature solid oxide fuel cell cathode. J Power Sources 2021;506:230230. https://doi.org/10.1016/j.jpowsour.2021.230230.

[87] Liping S, Na L, Qiang L, Lihua H, Hui Z. Preparation of NdBaCo2O5+δ–Ce0.9Gd0.1O1.95 composite cathode by in situ sol-mixing method and its high-temperature electrochemical properties. J Alloys Compd 2021:885. https://doi.org/10.1016/j.jallcom.2021.160901.

[88] Zhang Y, Shen L, Wang Y, Du Z, Zhang B, Ciucci F, et al. Enhanced oxygen reduction kinetics of IT-SOFC cathode with PrBaCo2O5+δ/Gd0.1Ce1.9O2-δcoherent interface. J Mater Chem A 2022;10:3495–505. https://doi.org/10.1039/d1ta09615j.

[89] Chiu TW, Lin MX, Shih HY, Hwang B yinn, Chang HY, Wang YM. Preparation and performance of PrLaNiO4 and (La0.75Sr0.2 Ba0.05)0.175Ce0.825O1.891 composite cathode material by solid state reaction for IT-SOFCs. Ceram Int 2017;43:S700–4. https://doi.org/10.1016/j.ceramint.2017.05.269.

[90] Wang Z, Sun H, Li J, Guo X, Hu Q, Yang Z, et al. Modified La0.6Sr0.4Co0.2Fe0.8O3-δ cathodes with the infiltration of Er0.4Bi1.6O3 for intermediate-temperature solid oxide fuel cells. Int J Hydrogen Energy 2021;46:22932–41. https://doi.org/10.1016/j.ijhydene.2021.04.121.

[91] Wang X, Afzal M, Deng H, Dong W, Wang B, Mi Y, et al. La0.1SrxCa0.9−xMnO3−Δ-Sm0.2Ce0.8O1.9 composite material for novel low temperature solid oxide fuel cells. Int J Hydrogen Energy 2017;42:17552–8. https://doi.org/10.1016/j.ijhydene.2017.05.158.

[92] Li P, Yang Q, Zhang H, Yao M, Yan F, Fu D. Effect of Fe, Ni and Zn dopants in La0·9Sr0·1CoO3 on the electrochemical performance of single-component solid oxide fuel cell. Int J Hydrogen Energy 2020;45:11802–13. https://doi.org/10.1016/j.ijhydene.2020.02.116.

[93] Huang J, Gao R, Mao Z, Feng J. Investigation of La2NiO4+δ-based cathodes for SDC-carbonate composite electrolyte intermediate temperature fuel cells. Int J Hydrogen Energy 2010;35:2657–62. https://doi.org/10.1016/j.ijhydene.2009.04.022.

[94] Caizán-Juanarena L, Zamudio-García J, Marrero-López D. Electrochemical investigation of Pr6O11 infiltration into La0.8Sr0.2MnO3-δ-Ce0.9Gd0.1O1.95 cathodes for IT-SOFC. Ceram Int 2023;49:33717–24. https://doi.org/10.1016/j.ceramint.2023.08.060.

[95] Jiang C, Zuo L, Liu Y, Zhang Z, Wang K, Zheng D, et al. F-doped LiNi0.8Co0.15Al0.05O2-δ cathodes with enhanced ORR catalytic activity for LT-SOFCs. J Alloys Compd 2023;940:168837. https://doi.org/10.1016/j.jallcom.2023.168837.

[96] Loureiro FJA, Macedo DA, Nascimento RM, Cesário MR, Grilo JPF, Yaremchenko AA, et al. Cathodic polarisation of composite LSCF-SDC IT-SOFC electrode synthesised by one-step microwave self-assisted combustion. J Eur Ceram Soc 2019;39:1846–53. https://doi.org/10.1016/j.jeurceramsoc.2019.01.013.

[97] Muhammed Ali SA, Anwar M, Raduwan NF, Muchtar A, Somalu MR. Optical, mechanical and electrical properties of LSCF–SDC composite cathode prepared by sol–gel assisted rotary evaporation technique. J Sol Gel Sci Technol 2018;86:493–504. https://doi.org/10.1007/s10971-018-4636-8.

[98] Ahuja A, Gautam M, Sinha A, Sharma J, Patro PK, Venkatasubramanian A. Effect of processing route on the properties of LSCF-based composite cathode for IT-SOFC. Bull Mater Sci 2020;43:1–9. https://doi.org/10.1007/s12034-020-2075-y.

[99] Yue Z, Jiang L, Ai N, Guan C, Jiang SP, Sun X, et al. Facile co-synthesis and utilization of ultrafine and highly active PrBa0.8Ca0.2Co2O5+δ-Gd0.2Ce0.8O1.9

composite cathodes for solid oxide fuel cells. Electrochim Acta 2022:403. https://doi.org/10.1016/j.electacta.2021.139673.
[100] Xu H, Dang L, Yan J, Wan F, Gong W. Preparation of a nano-size (La0.2Nd0.2Sm0.2Sr0.2Ba0.2)Co0.2Fe0.8O3-δ/SDC high-entropy oxide composite cathode. Mater Lett 2023;338:134029. https://doi.org/10.1016/j.matlet.2023.134029.
[101] Xi X, Huang L, Chen L, Liu W, Liu X, Luo JL, et al. Enhanced reaction kinetics of BCFZY-GDC-PrOx composite cathode for low-temperature solid oxide fuel cells. Electrochim Acta 2023;439:141617. https://doi.org/10.1016/j.electacta.2022.141617.
[102] Zhang X, Jin Y, Jiang Y, Zong X, Li Y, Xiong Y. Enhanced electrochemical performance and operating stability of La0.8Sr0.2Co0.2Fe0.8O3-δ fiber cathode via Gd0.2Ce0.8O1.9-δ coating for intermediate temperature solid oxide fuel cells. Appl Surf Sci 2023;630:157445. https://doi.org/10.1016/j.apsusc.2023.157445.
[103] Bai J, Han Z, Zhou D, Zhu X, Wang N, Chen R, et al. Preparation of Pr2NiO4+δ-La0.6Sr0.4CoO3-δ as a high-performance cathode material for SOFC by an impregnation method. Int J Hydrogen Energy 2023;48:6076–87. https://doi.org/10.1016/j.ijhydene.2022.11.123.
[104] Desta HG, Yang Y, Teketel BS, Yang Q, Song K, Zhu S, et al. Enhanced performance of La0.8Sr0.2FeO3-δ-Gd0.2Ce0.8O2-δ cathode for solid oxide fuel cells by surface modification with BaCO3 nanoparticles. Micromachines 2022;13. https://doi.org/10.3390/mi13060884.
[105] Rehman SU, Hassan MH, Batool SY, Kim HS, Song RH, Lim TH, et al. A highly stable Co3O4-GDC nanocomposite cathode for intermediate temperature solid oxide fuel cells. Int J Hydrogen Energy 2023. https://doi.org/10.1016/j.ijhydene.2023.03.341.
[106] Pikalova E, Bogdanovich N, Kolchugin A, Shubin K, Ermakova L, Eremeev N, et al. Development of composite LaNi0·6Fe0.4O3-δ-based air electrodes for solid oxide fuel cells with a thin-film bilayer electrolyte. Int J Hydrogen Energy 2021;46:16947–64. https://doi.org/10.1016/j.ijhydene.2021.02.217.
[107] Maguire E, Gharbage B, Marques FMB, Labrincha JA. Cathode materials for intermediate temperature SOFCs. Solid State Ionics 2000;127:329–35. https://doi.org/10.1016/S0167-2738(99)00286-6.
[108] Jun A, Kim J, Shin J, Kim G. Perovskite as a cathode material: a review of its role in solid-oxide fuel cell technology. Chemelectrochem 2016;3:511–30. https://doi.org/10.1002/celc.201500382.
[109] Lyskov NV, Galin MZ, Kostretsova NB, Eliseeva GM, Kolchina LM, Mazo GN. Electrochemical properties of composite cathode materials Pr1.95La0.05CuO4–Ce0.9Gd0.1O1.95 for intermediate temperature solid oxide fuel cells. Russ J Electrochem 2018;54:527–32. https://doi.org/10.1134/S1023193518060137.
[110] Seo J, Kim HW, Yu JH, Park HJ. Electrochemical properties of Ba0.5Sr0.5Co0.8Fe0.2O3 and BaZr0.65Ce0.20Y0.15O3 composite cathodes on Y-doped barium–cerium–zirconium oxide solid electrolyte. J Korean Ceram Soc 2022;59:217–23. https://doi.org/10.1007/s43207-021-00169-7.
[111] Küngas R, Bidrawn F, Mahmoud E, Vohs JM, Gorte RJ. Evidence of surface-reaction rate limitations in SOFC composite cathodes. Solid State Ionics 2012;225:146–50. https://doi.org/10.1016/j.ssi.2012.04.030.
[112] Venkataramana K, Madhuri C, Madhusudan C, Bhogi A, Srinivas B, Reddy CV. Influence of La3+, Sm3+and Dy3+ dopants on ceria solid electrolytes for IT–SOFCs. Mater Sci Semicond Process 2022;142:106495. https://doi.org/10.1016/j.mssp.2022.106495.
[113] Antoni L. Materials for solid oxide fuel cells : the challenge of their stability. Mater Sci Forum 2004;461–464:1073–90. https://doi.org/10.4028/www.scientific.net/msf.461-464.1073.
[114] Murray EP, Sever MJ, Barnett SA. Electrochemical performance of (La,Sr)(Co,Fe)O_3_-(Ce,Gd)O_3_ composite cathodes. Solid State Ionics 2002;148:27–34.
[115] Jafari M, Salamati H, Zhiani M, Shahsavari E. Enhancement of an IT-SOFC cathode by introducing YSZ: electrical and electrochemical properties of La 0.6 Ca 0.4 Fe 0.8 Ni 0.2 O 3-Δ -YSZ composites. Int J Hydrogen Energy 2019;44:1953–66. https://doi.org/10.1016/j.ijhydene.2018.10.151.
[116] Ivers-Tiffée E, Weber A, Herbstritt D. Materials and technologies for SOFC-components. J Eur Ceram Soc 2001;21:1805–11. https://doi.org/10.1016/S0955-2219(01)00120-0.
[117] Klein L, Aparicio M, Damay F. Sol-gel processing for battery and fuel cell applications. Handb Sol-Gel Sci Technol Process Charact Appl 2018;2573–93. https://doi.org/10.1007/978-3-319-32101-1_66.
[118] Javed MA, Muneer M, Abbas G, Shakir I, Batool R, Iqbal M, et al. Structural and electrochemical characterization of low-cost LixCu1-xCoy Fe1-yO3-δ cathode material for intermediate temperature solid oxide fuel cell. Ceram Int 2020;46:10348–55. https://doi.org/10.1016/j.ceramint.2020.01.031.
[119] Samreen A, Galvez-Sanchez M, Steinberger-Wilckens R, Arifin NA, Saher S, Ali S, et al. Electrochemical performance of novel NGCO-LSCF composite cathode for intermediate temperature solid oxide fuel cells. Int J Hydrogen Energy 2020;45:21714–21. https://doi.org/10.1016/j.ijhydene.2020.04.122.
[120] Padmasree KP, Lai KY, Manthiram A. Synthesis and characterization of Ca3-xLaxCo4-yCuyO9+δ cathodes for intermediate temperature solid oxide fuel cells. Ceram Int 2022;48:455–62. https://doi.org/10.1016/j.ceramint.2021.09.121.
[121] Li N, Wei B, Lü Z, Huang X, Su W. GdBaCo2O5+δ-Sm0.2Ce 0.8O1.9 composite cathodes for intermediate temperature SOFCs. J Alloys Compd 2011;509:3651–5. https://doi.org/10.1016/j.jallcom.2010.12.147.
[122] Filonova EA, Pikalova EY, Maksimchuk TY, Vylkov AI, Pikalov SM, Maignan A. Crystal structure and functional properties of Nd1.6Ca0.4Ni1-yCuyO4+δ as prospective cathode materials for intermediate temperature solid oxide fuel cells. Int J Hydrogen Energy 2021;46:17037–50. https://doi.org/10.1016/j.ijhydene.2020.10.243.
[123] Liu B, Zhang Y. Ba0.5Sr0.5Co0.8Fe0.2O3 nanopowders prepared by glycine-nitrate process for solid oxide fuel cell cathode. J Alloys Compd 2008;453:418–22. https://doi.org/10.1016/j.jallcom.2006.11.142.
[124] Anbo Y, Tian X, Liping S, Qiang L, Lihua H, Hui Z. Effects of rare earth doping on electrochemical properties of NdBaCo2O6-δ cathode materials. J Alloys Compd 2020;837. https://doi.org/10.1016/j.jallcom.2020.155563.
[125] Lu F, Yang M, Shi Y, Wu C, Jia X, He H, et al. Application of a negative thermal expansion oxide in SOFC cathode. Ceram Int 2021;47:1095–100. https://doi.org/10.1016/j.ceramint.2020.08.225.
[126] Geofrey Sahini M, Daud Lupyana S. Perspective and control of cation interdiffusion and interface reactions in solid oxide fuel cells (SOFCs). Mater Sci Eng B Solid-State Mater Adv Technol 2023;292:116415. https://doi.org/10.1016/j.mseb.2023.116415.
[127] Nie X, Chen Y, Mushtaq N, Rauf S, Wang B, Dong W, et al. The sintering temperature effect on electrochemical properties of Ce0.8Sm0.05Ca0.15O2-δ (SCDC)-La0.6Sr0.4Co0.2Fe0.8O3-δ (LSCF) heterostructure pellet. Nanoscale Res Lett 2019;14. https://doi.org/10.1186/s11671-019-2979-x.
[128] Zhou D, Fu J, Dong X, Yin Y mei, Wang Z, Xu J. A facial and effective way to prepare high-performance hetero-structured composite cathode for intermediate temperature solid oxide fuel cells. Int J Hydrogen Energy 2022;47:13112–20. https://doi.org/10.1016/j.ijhydene.2022.02.063.
[129] Li D, Jin Y, Liu C, Fu M, Zong X, Xiong Y. Effect of optimal GDC loading on long-term stability of La0.8Sr0.2Co0.2Fe0.8O3-δ/Gd0.2Ce0.8O1.9 composite cathodes. Ceram Int 2021;47:6591–6. https://doi.org/10.1016/j.ceramint.2020.10.248.
[130] Marshenya SN, Politov BV, Osinkin DA, Suntsov AY, Leonidov IA, Kozhevnikov VL. Advanced electrochemical properties of Pr0.9Y0.1BaCo1.8Ni0.2O6–δ – Ce0.8Sm0.2O1.9 composite as cathode material for IT–SOFCs. J Alloys Compd 2019;779:712–9. https://doi.org/10.1016/j.jallcom.2018.11.255.
[131] Wang X, Tang B, Wen P, Dong W, Wang L, Wang D. YSZ/LSM composite cathode deposited by solution precursor plasma spraying. Coatings 2022;12:10–48. https://doi.org/10.3390/coatings12030321.
[132] Aysal HE, Kılıç F, Çakmak G, Öztürk T. Thermal plasma synthesis of (La,Sr)CoO3-(La,Sr)2CoO4 composite cathodes for intermediate temperature solid oxide fuel cells (IT-SOFC). Int J Hydrogen Energy 2023. https://doi.org/10.1016/j.ijhydene.2023.07.031.
[133] Qi H, Qiu P, Zhang T, Liu D, Cheng M, Tu B. Self-assembled, high-performing cobalt-free Ba0.5A0.5Fe0.8Zr0.2O3-δ (A=Sr2+/Sm3+) composite cathode for intermediate-temperature solid oxide fuel cells. Ceram Int 2022;48:28669–77. https://doi.org/10.1016/j.ceramint.2022.06.181.
[134] Tan KH, Rahman HA, Azami MS, Yusop UA, Baharuddin NA, Ma'arof MIN. Electrochemical and material characteristics of Ba0.5Sr0.5Co0.8Fe0.2O3−δ-Sm0.2Ce0.8O1.9 carbonate perovskite cathode composite for low-temperature solid oxide fuel cell. Ceram Int 2022;48:34258–64. https://doi.org/10.1016/j.ceramint.2022.07.325.
[135] Li D, Zhang X, Liang C, Jin Y, Fu M, Yuan J, et al. Study on durability of novel core-shell-structured La0.8Sr0.2Co0.2Fe0.8O3-δ@Gd0.2Ce0.8O1.9 composite materials for solid oxide fuel cell cathodes. Int J Hydrogen Energy 2021;46:28221–31. https://doi.org/10.1016/j.ijhydene.2021.06.031.
[136] Melo KPV, Araújo AJM, Grilo JPF, Fagg DP, Macedo DA, Loureiro FJA. Understanding the oxygen reduction reaction of one-dimensional Ca3Co2O6 cathodes for SOFC. Int J Hydrogen Energy 2023. https://doi.org/10.1016/j.ijhydene.2023.05.186.
[137] Farhan S, Mohsin M, Raza AH, Anwar R, Ahmad B, Raza R. Co-doped cerium oxide Fe0.25xMnxCe0.75O2-δ as a composite cathode material for IT-SOFC. J Alloys Compd 2022;906:164319. https://doi.org/10.1016/j.jallcom.2022.164319.
[138] Fatah AFF, Mohamad AAA, Muchtar A, Hamid NAA. Physical characterization of LSCF-CuO via enhanced modified sol–gel method for intermediate temperature Solid oxide Fuel Cells (IT-SOFCs). Mater Today Proc 2020;46:2052–7. https://doi.org/10.1016/j.matpr.2021.03.141.
[139] Leng Y, Chan SH, Liu Q. Development of LSCF-GDC composite cathodes for low-temperature solid oxide fuel cells with thin film GDC electrolyte. Int J Hydrogen Energy 2008;33:3808–17. https://doi.org/10.1016/j.ijhydene.2008.04.034.
[140] Rosli AZ, Somalu MR, Osman N, Hamid NA. Physical characterization of LSCF-NiO as cathode material for intermediate temperature solid oxide fuel cell (IT-SOFCs). Mater Today Proc 2021;46:1895. https://doi.org/10.1016/j.matpr.2021.01.778. –900.
[141] Wang X, Fei M, Zhou C, Li W, Wang X, Shen X, et al. Enhanced proton conductivity and CO2-tolerance of intermediate-temperature protonic ceramic fuel cell with lanthanum tungstate-based composite cathode. Composites Part B 2023;253:110565. https://doi.org/10.1016/j.compositesb.2023.110565.
[142] Yousaf M, Akbar M, Hu E, Dong Y, Akhtar MN, Yousaf Shah MAK, et al. Tuning ORR electrocatalytic functionalities in CGFO-GDC composite cathode for low-temperature solid oxide fuel cells. Ceram Int 2022;49:1–9. https://doi.org/10.1016/j.ceramint.2022.10.095.
[143] Huang Z, Liu Z, Hu H, Wang J, Chen M, Cao B, et al. Evaluation of La0.6Sr0.4CoO3-δ-Ce0.85Sm0.075Nd0.075O2-δ composite cathodes for intermediate temperature solid oxide fuel cells. Ceram Int 2022;48:16319–25. https://doi.org/10.1016/j.ceramint.2022.02.182.
[144] Bhasin I, Mohan Kant K. Development of conducting Bi0.5Sr0.5FeO3–δ-Gd0.1Ce0.9O2–δ composite cathodes for IT-SOFCs. Mater Lett 2024;354:135344. https://doi.org/10.1016/j.matlet.2023.135344.
[145] Mushtaq N, Lu Y, Xia C, Dong W, Wang B, Shah MAKY, et al. Promoted electrocatalytic activity and ionic transport simultaneously in dual functional

Ba0.5Sr0.5Fe0.8Sb0.2O3-δ-Sm0.2Ce0.8O2-δ heterostructure. Appl Catal B Environ 2021;298:120503. https://doi.org/10.1016/j.apcatb.2021.120503.
[146] Deseure J, Bultel Y, Dessemond L, Siebert E. Theoretical optimisation of a SOFC composite cathode. Electrochim Acta 2005;50:2037–46. https://doi.org/10.1016/j.electacta.2004.09.012.
[147] Yamaguchi Y, Kagomiya I, Minami S, Shimada H, Sumi H, Ogura Y, et al. La0.65Ca0.35FeO3-δ as a novel Sr- and Co-free cathode material for solid oxide fuel cells. J Power Sources 2020;448:227426. https://doi.org/10.1016/j.jpowsour.2019.227426.
[148] Pelosato R, Cordaro G, Stucchi D, Cristiani C, Dotelli G. Cobalt based layered perovskites as cathode material for intermediate temperature Solid Oxide Fuel Cells: a brief review. J Power Sources 2015;298:46–67. https://doi.org/10.1016/j.jpowsour.2015.08.034.
[149] Fu C, Sun K, Zhang N, Chen X, Zhou D. Electrochemical characteristics of LSCF-SDC composite cathode for intermediate temperature SOFC. Electrochim Acta 2007;52:4589–94. https://doi.org/10.1016/j.electacta.2007.01.001.
[150] Zhou Q, Wang Y, Yang X, Bu F, Yang F, Wang M, et al. Basic properties of low thermal expansion coefficient (Y0.5Ca0.5)1−xInxBaCo3ZnO7+δ (x=0, 0.1, 0.2, 0.3) solid solutions for solid oxide fuel cell cathode materials. Mater Res Bull 2022;156:112001. https://doi.org/10.1016/j.materresbull.2022.112001.
[151] Lee SJ, Yong SM, Kim DS, Kim DK. Cobalt-free composite cathode for SOFCs: brownmillerite-type calcium ferrite and gadolinium-doped ceria. Int J Hydrogen Energy 2012;37:17217–24. https://doi.org/10.1016/j.ijhydene.2012.08.100.
[152] Zhang L, Liu M, Huang J, Song Z, Fu Y, Chang Y, et al. Improved thermal expansion and electrochemical performances of Ba 0.6Sr0.4Co0.9Nb0.1O 3-δ-Gd0.1Ce0.9O1.95 composite cathodes for IT-SOFCs. Int J Hydrogen Energy 2014;39:7972–9. https://doi.org/10.1016/j.ijhydene.2014.03.055.
[153] Ding X, Cui C, Guo L. Thermal expansion and electrochemical performance of La0.7Sr0.3CuO3-δ-Sm0.2Ce0.8O2-δ composite cathode for IT-SOFCs. J Alloys Compd 2009;481:845–50. https://doi.org/10.1016/j.jallcom.2009.03.132.
[154] Patro PK, Delahaye T, Bouyer E. Development of Pr0.58Sr0.4Fe0.8Co 0.2O3 -δ-GDC composite cathode for solid oxide fuel cell (SOFC) application. Solid State Ionics 2010;181:1378–86. https://doi.org/10.1016/j.ssi.2010.07.004.
[155] Lu Y, Mushtaq N, Yousaf Shah MAK, Irshad MS, Rauf S, Yousaf M, et al. Ba0.5Sr0.5Fe0.8Sb0.2O3-δ- Sm0.2Ce0.8O2-δ bulk heterostructure composite: a cobalt free Oxygen Reduction Electrocatalyst for low-temperature SOFCs. Int J Hydrogen Energy 2022;47:38348–60. https://doi.org/10.1016/j.ijhydene.2022.09.009.
[156] Wang SF, Hsu YF, Liao Y Le, Huang ST, Jasinski P. High-performance NdSrCo2O5+δ–Ce0.8Gd0.2O2-δ composite cathodes for electrolyte-supported microtubular solid oxide fuel cells. Int J Hydrogen Energy 2021;46:31778–87. https://doi.org/10.1016/j.ijhydene.2021.07.049.
[157] Zhu JH, Ghezel-Ayagh H. Cathode-side electrical contact and contact materials for solid oxide fuel cell stacking: a review. Int J Hydrogen Energy 2017;42:24278–300. https://doi.org/10.1016/j.ijhydene.2017.08.005.
[158] Agun L, Abd Rahman H, Ahmad S, Muchtar A. Durability and stability of LSCF composite cathode for intermediate-low temperature of solid oxide fuel cell (IT-LT SOFC): short review. Adv Mater Res 2014;893:732–7. https://doi.org/10.4028/www.scientific.net/AMR.893.732.
[159] Subhashini PVCK, Rajesh KVD. Manufacturing method of BSCF cathode for low-temperature solid oxide fuel cell-A review. Mater Today Proc 2022;62:2357–61. https://doi.org/10.1016/j.matpr.2022.04.653.
[160] Liu J, Jin F, Yang X, Niu B, Yu S, Li Y, et al. Synergistic electron doping and ion conductive phase incorporating of SrCoO3-δ as desirable cathode materials for intermediate-temperature solid oxide fuel cells. Ceram Int 2020;46:28332–41. https://doi.org/10.1016/j.ceramint.2020.07.337.
[161] Chen L, Chen F, Xia C. Direct synthesis of methane from CO2-H2O co-electrolysis in tubular solid oxide electrolysis cells. Energy Environ Sci 2014;7:4018–22. https://doi.org/10.1039/c4ee02786h.
[162] Yao C, Yang J, Zhang H, Chen S, Meng J, Cai K. Characterization of SrFe0.9-xCuxMo0.1O3-δ (x = 0, 0.1 and 0.2) as cathode for intermediate-temperature solid oxide fuel cells. Int J Energy Res 2021;45:5337–46. https://doi.org/10.1002/er.6156.
[163] Li Z, Wu T, Zhang Z, Sun H, Guo X, Hu Q, et al. Experimental study and first-principles calculation of enhanced electrochemical properties of Bi0.7Sr0.3FeO3-δ (BSFO) – La0.8Sr0.2MnO3-δ (LSM) composite cathodes for intermediate-temperature solid oxide fuel cells. J Power Sources 2024;591:1–8. https://doi.org/10.1016/j.jpowsour.2023.233898.
[164] Lee HY, Yu W, Lee YH. High-performance low-temperature solid oxide fuel cell with nanostructured lanthanum strontium cobaltite/yttria-stabilized zirconia cathode via advanced co-sputtering. J Alloys Compd 2024;972:172740. https://doi.org/10.1016/j.jallcom.2023.172740.
[165] Painter AS, Huang YL, Wachsman ED. Durability of (La0.8Sr0.2)0.95MnO3-Δ-(Er0.2Bi0.8)2O3 composite cathodes for low temperature SOFCs. J Power Sources 2017;360:391–8. https://doi.org/10.1016/j.jpowsour.2017.06.018.
[166] Zouhri K, Shinneeb M, Chikhalsouk M, Cress J. Solid oxide fuel cell cathode diffusion polarization: materials and exergy study. Energy Convers Manag 2021;231:113865. https://doi.org/10.1016/j.enconman.2021.113865.
[167] Chen Z, Wang X, Bhakhri V, Giuliani F, Atkinson A. Nanoindentation of porous bulk and thin films of La0.6Sr 0.4Co0.2Fe0.8O3-δ. Acta Mater 2013;61:5720–34. https://doi.org/10.1016/j.actamat.2013.06.016.
[168] Wang Q, Fan H, Xiao Y, Zhang Y. Applications and recent advances of rare earth in solid oxide fuel cells. J Rare Earths 2022. https://doi.org/10.1016/j.jre.2021.09.003.
[169] Li H, Su C, Wang C, Lü Z. Electrochemical performance evaluation of FeCo2O4 spinel composite cathode for solid oxide fuel cells. J Alloys Compd 2020;829:154493. https://doi.org/10.1016/j.jallcom.2020.154493.
[170] Idrees A, Jiang X, Jiang L, Zhang Q. Properties of composite cathodes composed of Pr3+-deficient perovskite oxide and ionic conductor Ce0.8Sm0.2O1.9. Ceram Int 2020;46:17532–9. https://doi.org/10.1016/j.ceramint.2020.04.051.
[171] Xia LN, You J, He ZP, Huang XW, Yu Y. Performances of nickel-doped SmBaCo2O5+δ-Sm0.2Ce0.8O1.9 composite cathodes for IT-SOFC. Int J Hydrogen Energy 2016;41:1176–86. https://doi.org/10.1016/j.ijhydene.2015.10.027.
[172] Zhang L, Li X, Zhang L, cai H, Xu J, Wang L, et al. Improved thermal expansion and electrochemical performance of La 0.4 Sr 0.6 Co 0.9 Sb 0.1 O 3-δ -Ce 0.8 Sm 0.2 O 1.9 composite cathode for IT-SOFCs. Solid State Sci 2019;91:126–32. https://doi.org/10.1016/j.solidstatesciences.2019.03.023.
[173] Guo W, Guo R, Cai G, Wu C, Liu Z. Preparation and characterization of PrBaCo1.9Ni0.1O5 + δ-Ce0.8Sm0.2O1.9 composite cathodes for intermediate temperature solid oxide fuel cells. Solid State Ionics 2015;283:10–5. https://doi.org/10.1016/j.ssi.2015.11.011.
[174] Chen M, Moon BH, Kim SH, Kim BH, Xu Q, Ahn BG. Characterization of La0.6Sr0.4Co0.2Fe 0.8O3-δ + La2NiO4+δ composite cathode materials for solid oxide fuel cells. Fuel Cell 2012;12:86–96. https://doi.org/10.1002/fuce.201100106.
[175] Jin F, Liu J, Shen Y, He T. Improved electrochemical performance and thermal expansion compatibility of LnBaCoFeO5+δ-Sm0.2Ce0.8O1.9 (Ln=Pr and Nd) composite cathodes for IT-SOFCs, 685. Elsevier Ltd; 2016. https://doi.org/10.1016/j.jallcom.2016.05.322.
[176] Chen Y, Yu T, Jin J, Zhang H. Triple perovskite Nd1.5Ba1.5CoFeMnO9−δ-Sm0.2Ce0.8O1.9 composite as cathodes for the intermediate temperature solid oxide fuel cells. Materials 2022;15:1–15. https://doi.org/10.3390/ma15103663.
[177] Solovyev AA, Shipilova AV, Rabotkin SV, Bogdanovich NM, Pikalova EY. Study of the efficiency of composite LaNi0.6Fe0.4O3-based cathodes in intermediate-temperature anode-supported SOFCs. Int J Hydrogen Energy 2023;48:22594–609. https://doi.org/10.1016/j.ijhydene.2023.02.011.
[178] Chen S, Jin J, Chen H, Guo L. Performance of YBa2Cu3O7-δ-La0.6Sr0.4Co0.2Fe0.8O3-δ composite cathode for intermediate-temperature solid oxide fuel cells. J Alloys Compd 2021;879:160458. https://doi.org/10.1016/j.jallcom.2021.160458.
[179] Aristizabal ME, Vega-Castillo J, Zimicz G, Prado F. Optimizing thermal expansion and polarization resistance of Ce1−xPrxO2−δ (x = 0.1 and 0.2)/La0.4Pr0.1Ba0.5Co0.975O3−δ composite cathode for IT-SOFC. Mater Lett 2023;350:2–5. https://doi.org/10.1016/j.matlet.2023.134966.
[180] Qin M. Study on preparation and properties of La0.6Sr0.4Co0.2Fe0.8O3-Δ-mGd0.2Ce0.8O2(M=0,30,40 and 50) composite cathode. E3S Web Conf 2023;406:1–5. https://doi.org/10.1051/e3sconf/202340601024.
[181] Liu C, Ma B, Lin Z, Zhou Y, Wu K. Exploring (1-x) CuFe2O4-xGd0.1Ce0.9O1.95 composites as the cathode materials for solid oxide fuel cells. Mater Lett 2022;325:132860. https://doi.org/10.1016/j.matlet.2022.132860.
[182] Desta HG, Tian D, Yang Q, Zhu S, Song K, Chen Y, et al. Developing a new Sr and Co-free composite cathode of solid oxide fuel cells with high performance. Chem Phys Lett 2022;806:140037. https://doi.org/10.1016/j.cplett.2022.140037.
[183] Solovyev AA, Shipilova AV, Rabotkin SV, Smolyanskiy EA, Shmakov AN. Magnetron sputtered LSC-GDC composite cathode interlayer for intermediate-temperature solid oxide fuel cells. Int J Hydrogen Energy 2022;47:37967–77. https://doi.org/10.1016/j.ijhydene.2022.08.281.
[184] Fu M, Lin X, Tan L, Zhang P, Xie H, Tao Z. Self-assembled Fe-doped PrBaCo2O5+δ composite cathodes with disorder transition region for intermediate-temperature solid oxide fuel cells. Int J Hydrogen Energy 2023;48:15229–37. https://doi.org/10.1016/j.ijhydene.2023.01.007.
[185] Li P, Wang Z, Huang X, Zhu L, Cao Z, Zhang Y, et al. Enhanced electrochemical performance of co-synthesized La2NiO4+δ-Ce0.55La0.45O2-δcomposite cathode for IT-SOFCs. J Alloys Compd 2017;705:105–11. https://doi.org/10.1016/j.jallcom.2017.02.131.
[186] Teng J, Xia T, Sun L, Huo L, Li Q, Zhao H. Designing one-step co-assembled Pr1.95Ce0.05CuO4-Ce0.8Pr0.2O1.9 composite cathode with extraordinary oxygen reduction activity for solid oxide fuel cells. J Alloys Compd 2022;929:167311. https://doi.org/10.1016/j.jallcom.2022.167311.
[187] Desta HG, Yang Q, Tian D, Zhu S, Lu X, Song K, et al. BaCO3 nanoparticles-modified composite cathode with improved electrochemical oxygen reduction kinetics for high-performing ceramic fuel cells. Catalysts 2022;12. https://doi.org/10.3390/catal12091046.
[188] Zhou J, Lin X, Long C, Li Y, Tao Z. Performance optimization for proton conducting SOFCs by impregnating triple-conducting NBSCF to form composite cathode. Mater Lett 2022;328:133181. https://doi.org/10.1016/j.matlet.2022.133181.
[189] Sun C, Shen Y, Wang F, Duan Q. Nanofibrous La0.7Sr0.3FeO3-δ-SDC cobalt-free composite cathode for proton-conducting solid oxide fuel cell. Mater Lett 2023;333:133687. https://doi.org/10.1016/j.matlet.2022.133687.
[190] Wu YC, Ye TY, Zhou Z. Study on microstructure and physical properties of Ba0.5Sr0.5Fe1−xCuxO3−δ–Ce0.8Sm0.15Ca0.05O1.875 cathode materials used in solid oxide fuel cell. Int J Hydrogen Energy 2023:1–10. https://doi.org/10.1016/j.ijhydene.2023.03.358.
[191] Sun C, Shen Y, Wang F, Duan Q. Optimization of a cobalt-free La0.7Sr0.3FeO3-δ-BaZr0.1Ce0.7Y0.2O3-δ composite cathode for proton-conducting solid oxide fuel cells. J Alloys Compd 2022;923:166447. https://doi.org/10.1016/j.jallcom.2022.166447.
[192] Li N, Sun L, Li Q, Xia T, Huo L, Zhao H. Novel and high-performance (La,Sr)MnO3 based composite cathodes for intermediate-temperature solid oxide fuel cells.

J Eur Ceram Soc 2023;43:5279–87. https://doi.org/10.1016/j.jeurceramsoc.2023.04.036.

[193] Fe L, Sr S, Cathode C, Fu M. Fabrication and study of LaNi0.6Fe0.4O3-δ and Sm0.5Sr0.5CoO3-δ composite cathode. Res Sq 2021:1–13. https://doi.org/10.21203/rs.3.rs-376930/v1License.

[194] Gao J, Li Q, Guo M, Sun L, Huo L, Zhao H. Improved electrochemical activity of Bi0.5Sr0.5FeO3-δ–Ce0.9Gd0.1O1.95composite cathode electrocatalyst for solid oxide fuel cells. Ceram Int 2021;47:748–54. https://doi.org/10.1016/j.ceramint.2020.08.185.

[195] Zhang H, Liu H, Cong Y, Yang W. Investigation of Sm0.5Sr0.5CoO3-δ/Co3O4 composite cathode for intermediate-temperature solid oxide fuel cells. J Power Sources 2008;185:129–35. https://doi.org/10.1016/j.jpowsour.2008.06.038.

Brain Structure and Function
https://doi.org/10.1007/s00429-024-02758-8

ORIGINAL ARTICLE

Altered DTI scalars in the hippocampus are associated with morphological and structural changes after traumatic brain injury

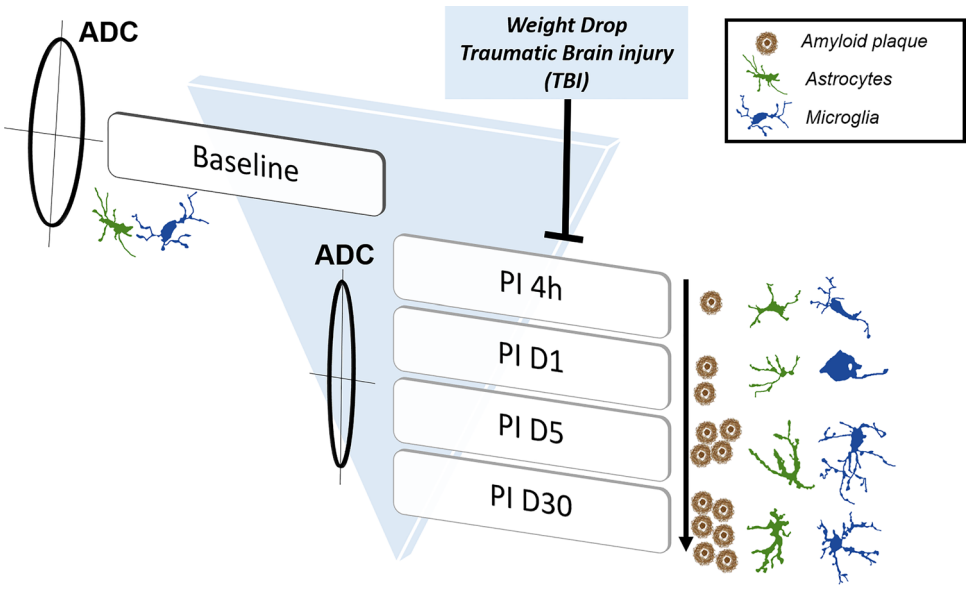
Palkin Arora[1,2] · Richa Trivedi[1] · Megha Kumari[1,3] · Kavita Singh[1] · Rajat Sandhir[2] · Maria M. D'Souza[1] · Poonam Rana[1]

Received: 30 March 2023 / Accepted: 3 January 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Blunt and diffuse injury is a highly prevalent form of traumatic brain injury (TBI) which can result in microstructural alterations in the brain. The blunt impact on the brain can affect the immediate contact region but can also affect the vulnerable regions like hippocampus, leading to functional impairment and long-lasting cognitive deficits. The hippocampus of the moderate weight drop injured male rats was longitudinally assessed for microstructural changes using in vivo MR imaging from 4 h to Day 30 post-injury (PI). The DTI analysis found a prominent decline in the apparent diffusion coefficient (ADC), radial diffusivity (RD), and axial diffusivity (AD) values after injury. The perturbed DTI scalars accompanied histological changes in the hippocampus, wherein both the microglia and astrocytes showed changes in the morphometric parameters at all timepoints. Along with this, the hippocampus showed presence of Aβ positive fibrils and neurite plaques after injury. Therefore, this study concludes that TBI can lead to a complex morphological, cellular, and structural alteration in the hippocampus which can be diagnosed using in vivo MR imaging techniques to prevent long-term functional deficits.

Graphical abstract



Keywords Diffusion tensor imaging · Microglia · Astrocytes · Aβ deposition · Closed head injury

Abbreviations

AD	Axial diffusivity
ADC	Apparent diffusion coefficient

Extended author information available on the last page of the article

Published online: 21 February 2024

CHI	Closed head injury
DAI	Diffuse axonal injury
DTI	Diffusion tensor imaging
FA	Fractional anisotropy
GFAP	Glial fibrillary acidic protein
IBA1	Ionized calcium binding adaptor molecule 1
PI	Post-injury
RD	Radial diffusivity
TBI	Traumatic brain injury

Introduction

Closed head injury caused by blunt impact from motor vehicle accidents, falls, and sports activities is a common form of TBI. The strong blunt and contact force is associated with long-term cognitive and motor deficits (Ng and Lee 2019). The functional deficits after trauma underlie a plethora of microstructural alterations. These structural impairments advance the major risk of dementia, Alzheimer's disease, Parkinson's disease, and stroke later in life (Graham and Sharp 2019). Early detection of cellular and structural integrity post-brain injury is essential for a complete diagnosis and progression of neuronal damage (Bramlett and Dietrich 2002).

Structurally, white matter abnormalities are common after TBI but studies have found quantitative structural and connectivity changes in gray matter regions like the hippocampus (Atkins 2011). Patients with moderate to severe TBI showed the development of mood disorders along with smaller hippocampal volume and were less likely to return to a productive life 1 year after trauma (Jorge et al. 2007). In another study, TBI-affected patients showed bilateral hippocampal atrophy mainly involving the hippocampal head (Ariza et al. 2006). Hippocampal damage is also associated with the induction of non-convulsive seizures after acute timepoint in moderate to severe TBI patients (Vespa et al. 2010). An impact to the surface of the brain can cause selective, bilateral hippocampus injury with associated abnormalities in dentate gyrus physiology with a patterned cell loss similar to pathology of temporal lobe epilepsy (Lowenstein et al. 1992). Therefore, clinically hippocampus is observed as a vulnerable brain region and anatomic damage can lead to life-long impairment post-TBI.

In addition, hippocampus-associated microstructural alterations have been observed in the pre-clinical models of TBI. Modifications in DTI scalars like FA, AD, and RD were observed in both gray and white matter regions (Sierra et al. 2015; Mohamed et al. 2020). In a study involving severe closed head TBI, DTI showed reduced FA and RD values in the gray matter regions including, the hippocampus post 14 days of injury (Mohamed et al. 2020). In another study, gray matter reorganization was observed in rat hippocampus using DTI which showed increased FA with a decrease in RD values in the CA3 region of lateral fluid percussed injured animals (Sierra et al. 2015), thus indicating structural derangement in the hippocampus and associated microstructural damage as important determinant of TBI pathology.

Blunt impact and compression can lead to dynamic deformation putting axonal tracts at risk of shearing forces and diffuse axonal injury (DAI) (Su and Bell 2016). Axonal injury can further lead to cytoskeletal compaction, and axonal swellings with the deposition of amyloid plaques (Siedler et al. 2014). Injury can simultaneously alter the surveillance behavior of the astrocytes and microglia, promoting hypertrophy and reactive glial population. This response can be both beneficial and detrimental, as astrocytes promote glial scarring at lesion sites but can also induce neurotoxicity by activating the microglial population to their reactive state (Liddelow et al. 2017; Bennett and Viaene 2021). Importantly, increased cellularity and inflammation in the brain can also elevate cohesiveness and change diffusivity parameters. It has been observed that glial scarring and gliosis contribute to microscopic structural changes in the brain leading to changes in diffusion anisotropy and tractography after TBI (Budde et al. 2011; Mira et al. 2021).

Therefore, to assess longitudinal changes in the hippocampal region after moderate CHI, rats were subjected to in vivo imaging at 4 h, Day 1, Day 5, and Day 30 PI. DTI scalars were measured at all timepoints to check the progression of injury-induced transformation. Further to evaluate the microstructural alterations and understand the crucial role of the glial population in mediating secondary injury response, morphological studies were performed on both astrocytic and microglial populations. Finally, to assess the axonal pathology, diffuse axonal injury (DAI) was determined.

Material and methods

Animals and injury protocol

Total 36 male Sprague Dawley (SD) rats weighing between 200 and 250 g (male, 6–8 weeks of age) were recruited from the experimental animal facility (EAF), INMAS, New Delhi. Rats were kept in cages at an optimum temperature and given standard pellet diet and water. After 7 days of acclimatization, experiments were conducted on the rats.

To replicate closed head blunt injury, Modified Marmarou's weight drop model was used (Arora et al. 2022), which consisted of a brass rod of 450 g inside an acrylic tube of 1 m. For moderate head injury, the brass rod was allowed to free fall from a height of 50 cm above the sagittal midway of the rats' brains. The rats were anesthetized using intraperitoneal administration of ketamine (80 mg/kg body weight)/

xylazine (10 mg/kg body weight) mixture and placed on an acrylic box filled with the foam bed. 30–40% mortality was observed within 4 h of injury to the rats. The experimental protocols were approved by the EAF, INMAS, DRDO (8/GO/RBi/S/99/CPCSEA/INM/IAEC/2021/02) and were in accordance with the guidelines for humane use and care of experimental animals. Exclusion criteria included any sign of illness during the start of the study.

Study paradigm

Animals were divided into two groups as described in Fig. 1—(a) longitudinal MRI and DTI studies, 10 rats were used for in vivo imaging studies on Day 0 (Baseline) and PI 4 h, Day 1, Day 5, and Day 30; (b) for histological studies, rest of the 15 rats were grouped into control, PI 4 h, PI Day 1, PI Day 5, and PI Day 30 groups ($n = 3$ each).

The control/baseline rats in both the study were anaesthetized and placed on the foam bed but they did not receive any injury.

MRI protocol

MR imaging was done using a Bruker Biospec 7.0 Tesla (Avance III) 30-cm horizontal bore magnet (Bruker Biospin, Ettlingen, Germany). The imager consisted of.

1. Receiver: an array coil of four channel for rat brain
2. Transmitter: a linear coil.

After anesthesia, rats were placed on a rat bed and RF array coil was kept to cover the whole head of the rat.

A fast low angle single shot (FLASH) sequence was obtained to localize the rat brain and a three-slice scout image was obtained (axial, mid-sagittal, and coronal).

A rapid acquisition with relaxation enhancement (RARE) sequence was used to acquire T2-weighted images of the control and injured rat brain using following parameters.

Field of view, 4*4 cm; matrix size, 256 × 256; repetition time/ echo time (TR)/(TE) = 2000/13 ms; slice thickness, 1 mm; number of slices = 15.

A multiple-shot spin echo EPI sequence was used to acquire DTI images with the following parameters:

Field of view, 4 cm × 4 cm; matrix size, 128 × 128; TR/TE, 5000 ms/34.46 ms; number of gradient encoding directions, 46; width** b, 672 s mm^{-2}; slice thickness, 1 mm; number of slices, 15; number of segments, 4; number of averages = 1; total imaging time = 12 min 55 s.

DTI data processing

The DTI data obtained from all timepoints were processed as described earlier (Trivedi et al. 2012). After image cropping and correcting of distortions, isotropic voxels were obtained and decoded to get the tensor field for each voxel. The tensor field data were then diagonalized using the analytic diagonalization method to obtain the eigenvalues (λ1, λ 2, and λ3) and the three orthonormal eigenvectors. The eigenvalues were used to compute the DTI metrics such as the apparent diffusion coefficient (ADC), fractional anisotropy (FA), radial diffusivity (RD), and axial diffusivity (AD) for each voxel. Regional FA, ADC, RD, and AD values were obtained by placing the bilateral regions of interest (ROIs; Fig. 2) on the hippocampus. The size of ROIs and slice were kept constant for the analysis between each rat. The hippocampus sub-fields were not analyzed due to obscured distinction between these areas because of lower resolution and contrast.

Immunohistochemistry

After transcardial perfusion with phosphate buffer saline (PBS 0.1 M, pH 7.4) and 4% (w/v) paraformaldehyde in 0.1 M PBS (pH 7.4), rat brains were post-fixed in buffered formalin. Coronal sections (5 µm) mounted on a charged slide using a microtome were then processed for immunohistochemistry. The brain slices were first defatted using xylene and rehydrated in 100% to 30% ethanol:water mixture. Antigen retrieval was performed using sodium citrate buffer at 80 °C for 1 h. Then Mouse and Rabbit Specific HRP/DAB (ABC) Detection IHC kit (#ab64264) was used where a 15 min H_2O_2 treatment was done followed by a protein block from the kit. The sections were then incubated with respective primary antibodies anti-IBA1 (#ab178846, RRID: AB_2636859; 1:100), anti-GFAP (#MA5-12,023, RRID: AB_10984338; 1:100) and anti-Aβ (# 51-2700, RRID: AB_87659; 1:200) overnight at 4 °C (Table 1). Secondary antibody (biotin-conjugated) was applied to sections and kept for 10 min. Following this, the sections were incubated with avidin-peroxidase complex. The brain slices were

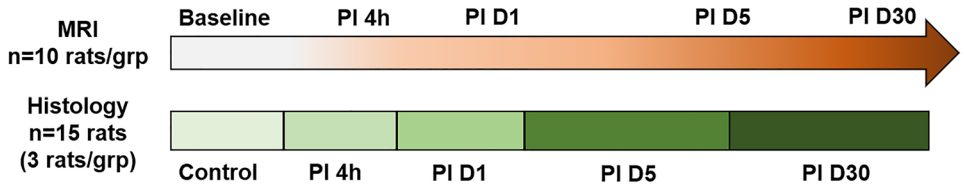


Fig. 1 Schematic representation of the study paradigm

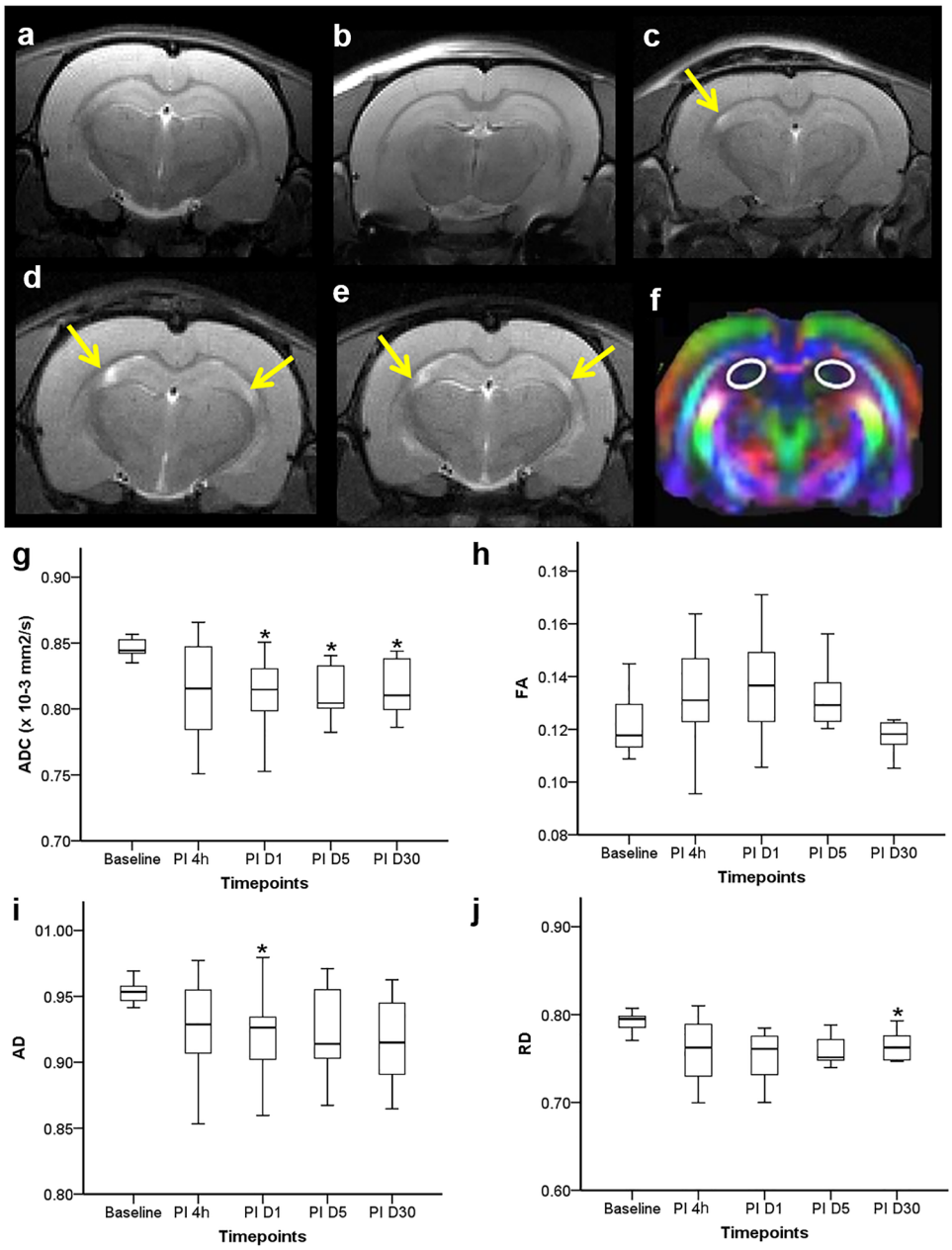


Fig. 2 In vivo imaging data at all timepoints PI; T2-weighted image of (**a**) baseline, (**b**) PI 4 h, (**c**) PI Day 1, (**d**) PI Day 5, (**e**) PI Day 30, (**f**) FA map of baseline. Boxplot of DTI scalars at all timepoints after injury (**g**) ADC ($\times 10^{-3}$ mm^2/s), (**h**) FA, (**i**) AD ($\times 10^{-3}$ mm^2/s), (**j**) RD ($\times 10^{-3}$ mm^2/s). Yellow arrows show presence of hyperintensity in the hippocampus. In the boxplots, the whiskers represent the minimum and the maximum value, the limits of the box represent the first and the third quartile, the median is depicted by the line that divides the box into two parts. *significantly different from control group ($^*p < 0.05$)

Table 1 List of antibodies used with their specifications

Antibodies	Immunogen	Host species/type	Catalog number /RRID	Concentration used
Anti-IBA1	Synthetic peptide	Rabbit monoclonal	Abcam, #ab178846/ RRID: AB_2636859	1:100
Anti- GFAP	Glial fibrillary acidic protein	Mouse monoclonal	Invitrogen, #MA5-12,023, RRID: AB_10984338	1:100
Anti-Aβ	A 22 amino acid synthetic peptide derived from the C-terminus of the human β-amyloid precursor protein (β-APP)	Rabbit polyclonal	Invitrogen, # 51-2700, RRID: AB_87659	1:200

then stained with 3,3′-diaminobenzidine (DAB) solution for 3–5 min and examined under the microscope (LMI BM-Prime, London, England) (Arora et al. 2022).

On the other hand, the anti-Aβ sections were incubated with Goat Anti-Rabbit IgG H&L (Alexa Fluor® 488) (# ab150077, RRID: AB_2630356; 1:200). To observe neurite and amyloid plaques, Thioflavin S staining was performed by incubating the defatted and rehydrated slides in 1% Thioflavin S (#1326-12-1) for 1 h and observed under fluorescence microscope (LMI BM-Prime, London, England).

Image analysis

The 40X photomicrographs obtained after IBA1 and GFAP staining was processed using FIJI (NIH, Bethesda, MD, USA, RRID: SCR_002285) while keeping the area of analysis (hippocampus) common for each photomicrograph. The section used for morphological analysis was kept similar to the MR slice on which DTI analysis was performed. The number of astrocytes and microglial cells was counted manually while the morphological analysis was performed using the *fraclac* plugin as described earlier (Arora et al. 2022). The morphological analysis was performed to observe time-dependent alterations in the astrocytic and microglial shape and structure after injury. Both the GFAP-positive and IBA1-positive photomicrographs of hippocampal region were processed in the same manner. The original histology images were converted to 8-bit images and then gray scaled using default threshold settings. Randomly single cells were selected from the threshold images (three cells/photomicrograph, three sections/ group) and then binary and outline images were obtained using FIJI. The outlined microglial and astrocytes were then analyzed using the *fraclac* plugin using four grid positions by the box counting method. The parameters included fractal dimension, lacunarity (heterogeneity), density (cell shape), span ratio (cell size), and circularity for both astrocytes and microglial population (Pirici et al. 2009; Young and Morrison 2018). All the data were observed by two viewers who were blinded to the groups of the study.

Statistical analysis

All values are expressed as mean ± standard deviation (SD). The longitudinal in vivo imaging data were analyzed using repeated-measures analysis of variance (ANOVA) ($n = 10$) and the histology data was analyzed using one-way ANOVA ($n = 3$). The post hoc analysis was performed using Bonferroni correction for comparison between all timepoints in IBM SPSS statistical software version 20 (IBM Corp., Armonk, N.Y., USA). Values with $p < 0.05$ were considered statistically significant.

Results

Longitudinal lesion characteristics on T2-weighted MRI after injury

Conventional T2-weighted images were observed for progression of injury-induced changes. Out of ten rats, four rats appeared normal on T2-weighted images at all the timepoints, while rest of the six animals showed lesions at the site of the injury as early as 4 h PI. PI Day 1, these six rats showed poorly demarcated hyperintensities in both hippocampi which persisted till PI Day 5 and Day 30. The hyperintensity observed was due edema, with no midline shift or mass effect. The other parenchymal structures also appeared typical of their normal appearance. (Fig. 2).

Longitudinal changes in the diffusivity indices after injury

The hippocampus showed microstructural alterations after moderate injury which was evident with modified DTI scalars. The overall diffusivity of the hippocampus denoted by ADC was found to be significantly decreased from Day 1 to Day 30 ($p < 0.05$) PI. The coefficient of diffusion along the long axis of ellipsoid AD showed a decline at Day 1 ($p < 0.05$) after injury. RD, the coefficient perpendicular to the long axis, decreased significantly only at Day 30 PI ($p < 0.05$) (Fig. 2).

Temporal morphological changes in the astrocytic and microglial population after injury

Microglia

The number of microglial (IBA1 positive) cells per field was found to be increased on PI Day 30 ($p < 0.001$). The morphological parameters analyzed using the *fraclac* plugin were also observed to be altered after injury. The fractal dimension significantly increased on PI Day 5 ($p < 0.01$) and PI Day 30 ($p < 0.001$). The heterogeneity of microglial processes denoted by lacunarity was significantly increased on PI Day 30 ($p < 0.001$). The cell size denoted as density was increased at PI Day 1 ($p < 0.001$) while decreased at PI Day 30 ($p < 0.001$). The cell elongation denoted by span ratio also increased on PI 4 h ($p < 0.001$), PI Day 1 ($p < 0.001$) but decreased on PI Day 5 ($p < 0.001$). The circularity (cell outline) of microglia was decreased on PI Day 1 ($p < 0.05$) but increased on PI Day 30 ($p < 0.001$) (Fig. 3).

Astrocytes

The astrocytic (GFAP positive) population per field was found to be increased on PI Day 1 ($p < 0.05$) and PI

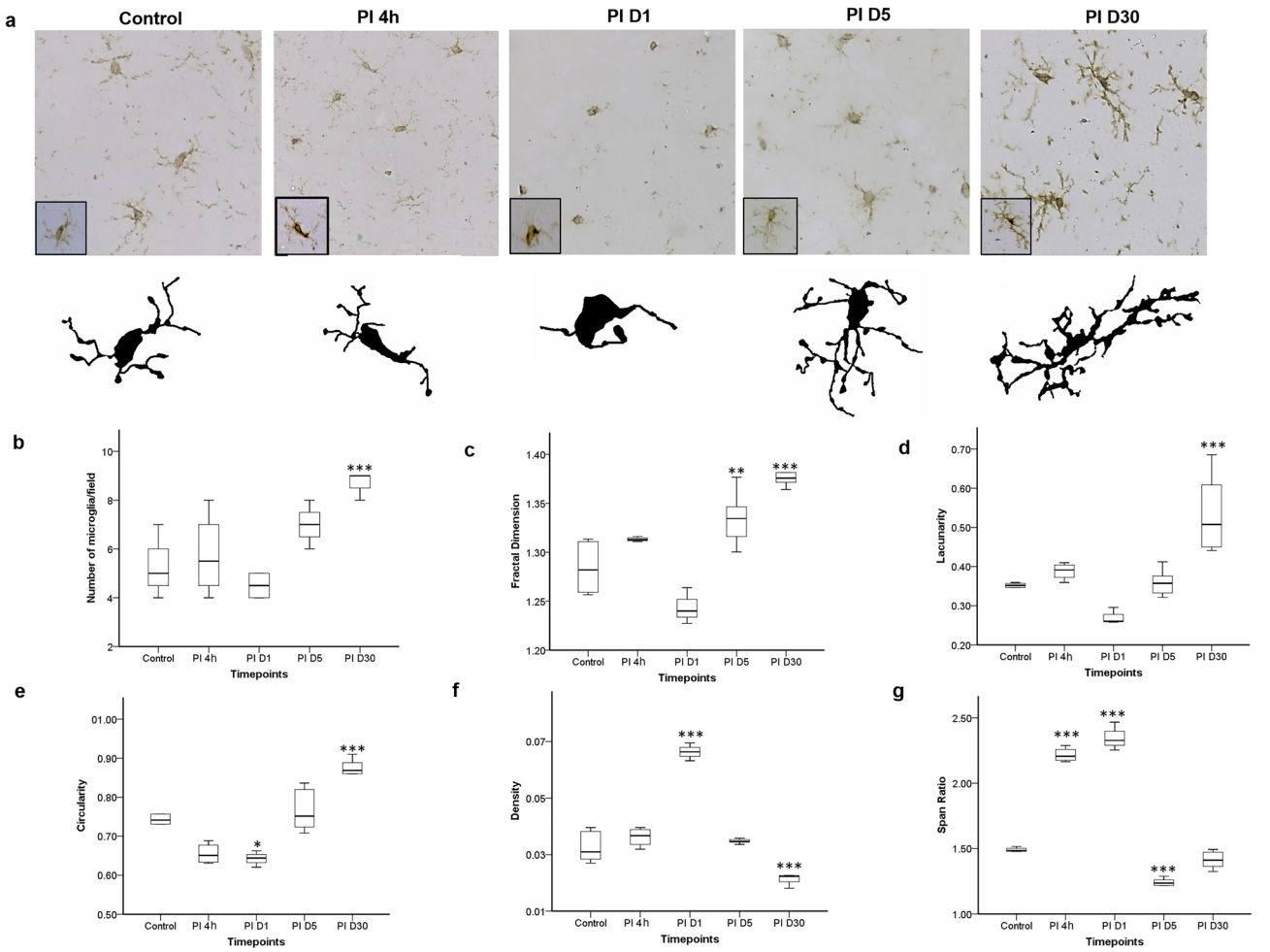


Fig. 3 Microglial morphological analysis at all timepoints after injury. (**a**) Photomicrographs of IBA1-stained sections, the inlet shows the single microglial cell and their corresponding binary images below. (**b**) Boxplot showing number of microglial per field, boxplot of morphological parameters quantified using *fraclac* plugin ($n = 10$–12 cells); (**c**) fractal dimension, (**d**) lacunarity, (**e**) circularity, (**f**) density, (**g**) span ratio. In the boxplots, the whiskers represent the minimum and the maximum value, the limits of the box represent the first and the third quartile, the median is depicted by the line that divides the box into two parts.*,**,***significantly different from control group ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$). Scale: 100 µm; magnification: 40X

Day 30 ($p < 0.05$). The morphological parameters analyzed using the *fraclac* plugin showed a variable astrocyte form at different timepoints. The fractal dimension was decreased on PI Day 1 but increased on PI Day 30 but these alterations were not significant. The lacunarity was significantly increased on PI Day 1 ($p < 0.05$). The density was decreased at PI 4 h ($p < 0.05$), PI Day 1 ($p < 0.001$), and PI Day 5 ($p < 0.01$). The span ratio increased on PI Day 1 ($p < 0.001$). The circularity of astrocytes was increased on PI Day 30 ($p < 0.05$) (Fig. 4).

A schematic representation of microglia and astrocytic population with respect to the change in their fractal dimension and lacunarity shows dynamic nature of both the glial population after injury (Fig. 5).

Temporal profile of axonal pathology after injury

Axonal pathology was assessed using Thioflavin S staining which showed the presence of neurite plaques on PI Day 5 and PI Day 30. While the presence of Aβ positive fibrils started appearing as early as PI 4 h and increased till PI Day 30 (Fig. 6).

Discussion

TBI assessment in a clinical setting involves a CT scan which helps to identify skull fractures and vascular injuries, but sensitive techniques including MR-based DTI can assist

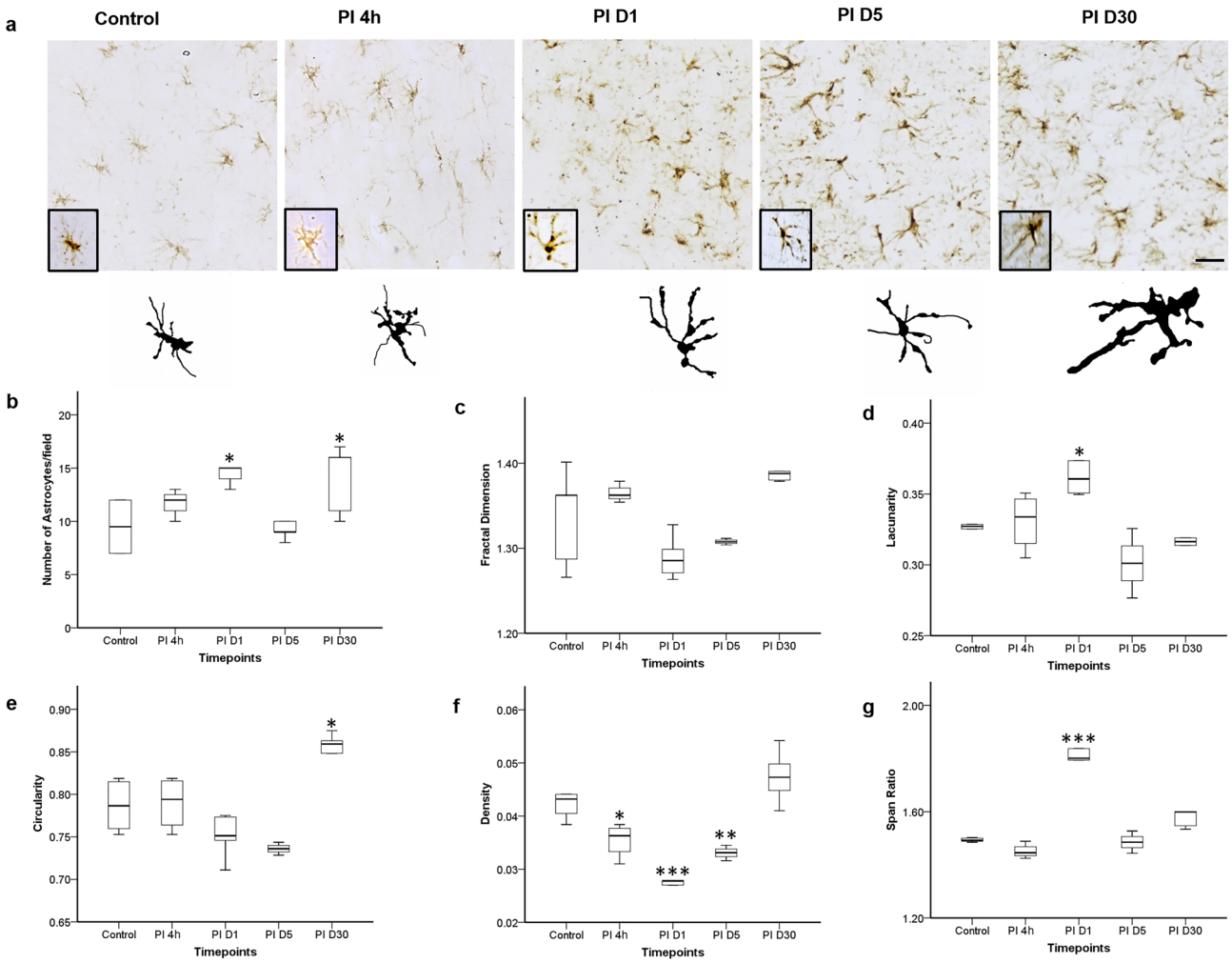


Fig. 4 Astrocytes morphological analysis at all timepoints after injury. (**a**) Photomicrographs of GFAP-stained sections, the inlet shows the single astrocytes and their corresponding binary images below. (**b**) Boxplot showing number of astrocytes per field, boxplot of morphological parameters quantified using *fraclac* plugin ($n = 10$–12 cells); (**c**) fractal dimension, (**d**) lacunarity, (**e**) circularity, (**f**) density, (**g**) span ratio. In the boxplots, the whiskers represent the minimum and the maximum value, the limits of the box represent the first and the third quartile, the median is depicted by the line that divides the box into two parts. *,**,***significantly different from control group ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$). Scale: 100 µm; magnification: 40X

in the detection of DAI, a hallmark of CHI (Hulkower et al. 2013). The present study found a decrease in ADC values from PI Day 1 to PI Day 30, with a decrease in AD at PI Day 1 and RD at PI Day 30 in the hippocampus. Along with this, a temporal morphological change was observed in astrocytes and microglia with presence of Aβ deposition.

DTI and specially ADC maps help in the identification of cytotoxic and vascular edema. Cytotoxic edema is caused during excitotoxic and ischemic brain injury which leads to accumulation of fluid in the neurons and glial cells (Ho et al. 2012). The gray matter region is more susceptible to cytotoxic edema because of higher astrocyte density and metabolic activity. This leads to restriction in the diffusion of water molecules leading to lower signal on ADC images. While secondary to cytotoxicity and repetitive insults to ion pumps and breakdown of blood–brain barrier, it can precede into vasogenic edema (Ho et al. 2012). TBI-induced edema is a combination of both vascular and cytotoxic edema, where MRI studies have shown that cytotoxic edema plays a more significant role than vasogenic edema (Winkler et al. 2016). Recently, presence of edema post-injury has also been observed due to alteration in the glymphatic system wherein reduced activity of astrocytic aquaporin-4 channel leads to interstitial fluid accumulation and decreased waste clearance from the brain (Ferrara et al. 2022).

In the present study, decreased ADC values were observed from Day 1 till Day 30 PI. This decrease in the overall diffusivity in the hippocampus can be an outcome of cytotoxic edema where increased signal in T2-weighted images also confirmed the presence of edematous hippocampus

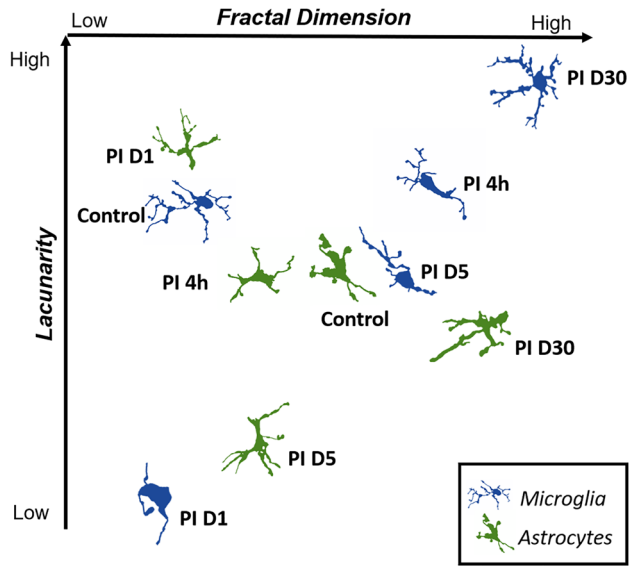


Fig. 5 Schematic representation of time-dependent changes in the fractal dimension and lacunarity of the microglial and astrocytic population from the control group

from Day 1 to Day 30 PI. In another DTI study, a significant upregulation in ADC values was seen after 60 min of injury which reached minimal levels at Day 7–14 (Barzó et al. 1997). While a study on humans observed a decrease in ADC values till sub-acute stage after injury (Newcombe et al. 2008)**.** Importantly, a dynamic relation exists between the type and degree of edema after TBI; wherein after cerebral contusion, it has been observed that non-damaged contralateral side has predominance of cytotoxic edema as compared to the damaged site. Similarly, it was found that contusion penumbra of CCI induced rats showed swollen gliocytes in electron microscopy with decreased ADC values after 7 days post-injury (Ren and Lu 2019). The decreased ADC values were accompanied by decline in the diffusion of water molecules in the parallel (AD) and perpendicular (RD) direction of the ellipsoid at PI Day 1 and PI Day 30, respectively. The changes in DTI scalars observed are similar to a previous study on assessment of microstructural alterations after diffuse TBI, where reduction in RD and AD values was found in the hippocampus (Mohamed et al. 2020). While another study on CCI model has found a decrease in AD and RD value in the cortex due to elevation in reactive astrocytes highlighting glial scarring as an important pathological change that can affect the diffusion metrics (Budde et al. 2011).

The glial cells play a crucial role in mediating inflammatory response, cellular damage, and edema after injury (Karve et al. 2016; Mira et al. 2021). Reactive astrogliosis is a hallmark of TBI and in response to diffuse injury, the dynamic strain to axons can lead to cytoskeletal disassembly without any gross destruction. Along with this, astrocytic mechanosensitive ion channels can lead to influx of Na^+ ions, resulting in cellular edema (Burda et al. 2016). In the present study, an increase in astrocytic population was observed in the hippocampus as early as PI Day 1 and Day 30, while microglial population was elevated at PI Day 30. A surge in the glial cells increased the cellularity which can also affect the diffusion of water molecules in the hippocampal region. Studies have found that glial cellularity can affect the anisotropy of the brain which can be detected using DTI; but how the complex temporal changes in the glial morphology results in microstructure derangement and DTI indices is still not known completely (Budde et al. 2011). A recent study has observed that microglial polarization is correlated with myelination/demyelination process via release of immune mediators and cytokines which further results in altered DTI scalars in the spinal cord (Chu et al. 2019). Another study on juvenile mild TBI has observed that distinct astrocytic morphology is correlated with long-term

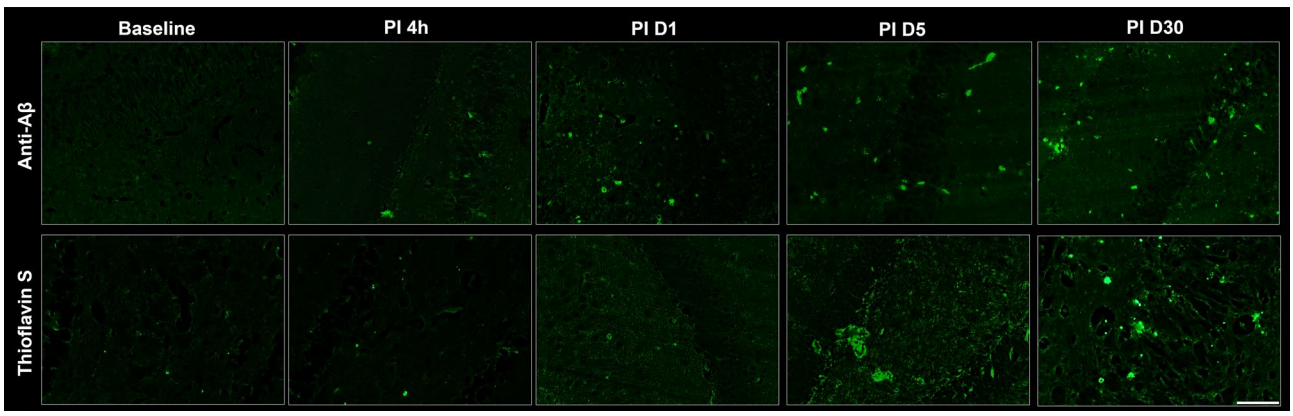


Fig. 6 The panel shows anti-Aβ and Thioflavin-S-stained sections at all timepoints PI. The appearance of green fluorescent Aβ fibrils and plaques increases with time after injury. Scale: 100 µm; magnification: 20X

structural changes and altered DTI scalars (Clément et al. 2020). Therefore, in the present study, extensive morphological changes in astrocytes and microglial were quantified to observe temporal changes in the hippocampus (site distant to impact) after TBI.

The astrocytic shape and size with its processes transformed at all timepoints. The astrocytes cell size (density) was reduced at PI 4 h, Day 1, and Day 30. The reduction in astrocytic size could indicate atrophy of astrocytic branches as observed in other neurological disorders like *Status Epilepticus* (Plata et al. 2018). The astrocytes were elongated with increased lacunarity at PI Day 1, indicating altered ramification due to outgrowth of particularly long processes leading to notable rotational variance (lacunarity). The cellular outline also started to increase at chronic timepoint with lacunarity reaching normal levels as compared to control (Karperien and Jelinek 2015; Schiweck et al. 2018). This concludes that diffuse brain injury does not lead to drastic alteration in the astrocytic morphology but can affect the cell size and the astrocytic process length.

On the other hand, microglial population showed enlarged cell size and elongated span ratio with decreased circularity at PI 4 h and Day 1, clearly suggesting activated M1 morphology. While at PI Day 5 and Day 30, the cell size decreased with short span ratio and increased fractal dimension. There was also a significant increase in lacunarity PI Day 30. The morphology attained by the microglial population from PI Day 5 to Day 30 was complex, heterogenous, rod-like, and ramified. The rod-like microglia have been earlier reported after TBI but also with the presence of train-like structures which were absent in our study (Taylor et al. 2014; Witcher et al. 2018). Thus, our study observed that microglial had activated morphology during acute phase which can be neurotoxic with M1 polarization potential as reported earlier (Kumar et al. 2022). Progressing into subacute and chronic stage, the microglial cells had increased ramifications with decreased shorter cells reflective of primed M2 cell population that have morphology different from healthy microglial population. This rod-like appearance of microglial can also have both M1/M2 phenotypes, as according to recent literature, their function is not clear and the tools to study function of these microglial population are lacking (Giordano et al. 2021). Therefore, the alterations in DTI scalars could be due to volumetric and cellular changes in the hippocampus; a consequence of transformed glial morphology after blunt impact (Haber et al. 2017; Lee et al. 2020).

Brain injury-induced axotomy is an important determinant of axonal injury and neuro phenotypic response. Axotomy involves modification in the key cytoskeletal elements leading to altered axonal transport and neurotransmission (Siedler et al. 2014). The presence of neurite plaques in thioflavin S staining at Day 5 and Day 30 PI marked axonal pathology in the hippocampus. While Aβ fibrils were seen from 4 h till Day 30 PI. Supporting, earlier studies using electron microscopy have depicted a progressive axonal pathology occurring for months after TBI, along with neuroinflammatory response and glial activation (Rodriguez-Paez et al. 2005). This deposition of Aβ from acute timepoint can affect the axonal function, and its accumulation in the chronic stages after injury can result in impaired functional recovery and onset of neurodegeneration (Johnson et al. 2010). DAI-induced microstructural changes can also alter the diffusion of water molecules leading to decreased diffusion anisotropy but no definite trend for DTI as an identification biomarker has been established (Arfanakis et al. 2002).

Conclusively, the intracellular or cytotoxic edema with alterations in glial morphology and DAI act as the confounding factors in hippocampus that can lead to microstructural alterations. Since hippocampal region is distant from site of contact injury, the ADC values were decreased suggesting that edema along with glial response can progress toward non-damaged area and can lead to severe pathological alterations. The astrocytes and microglial activation can further lead to cytotoxic effects including AD-like pathology after TBI.

Conclusion

CHI in rats leads to progressive structural and morphological changes which can be diagnosed using in vivo imaging techniques including DTI. T2-weighted images showed presence of hyperintensity in the hippocampus PI. The longitudinal assessment of DTI scalars showed a persistent decrease in ADC values in the hippocampus of injured rats from PI Day 1. The AD and RD were found to be decreased at PI Day 1 and Day 30, respectively. This modification in the DTI scalars was due to increased cellularity and morphological transformation of the astrocytic and microglial population. Both the glial population showed a kaleidoscope of shapes and forms at all timepoints after injury. Along with the morphological and structural modification in the hippocampus, Aβ deposition and neurite plaques were evident after injury. Thus, this study highlights an important contribution of cellular and structural transformation that occurs after injury and can be a contributing factor for altered DTI scalars.

Limitations to the study

The study was conducted in male rats, and therefore lack injury-induced alterations in the female rats. Since microglial population shows sexual dimorphism, considering female rats in further studies will be essential.

Author contributions All authors had access to the data and they take responsibility of the integrity and accuracy of the data. *Conceptualization, methodology and analysis* R.T., P.A.; *data curation and original draft preparation* P.A, R.T; *sample preparation and data acquisition* P.A., K.S, M.K.; *formal analysis* K.S, P.A.; *visualization, reviewing, editing* R.T., K.S., M.D., P.R., R.S. All authors have read and approved the final version of the manuscript.

Funding This work was supported by the Defence Research & Development Organisation (DRDO), Ministry of Defence, India (INM-324). Palkin Arora is supported by DRDO fellowship (EST/3009/JRF/2020).

Data accessibility statement The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declare no conflict of interest.

Research involving human and/or animal rights All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. This article does not contain any studies with human participants performed by any of the authors.

References

Arfanakis K, Haughton VM, Carew JD et al (2002) Diffusion tensor MR imaging in diffuse axonal injury. AJNR Am J Neuroradiol 23:794

Ariza M, Serra-Grabulosa JM, Junqué C, Ramírez B, Mataró M, Poca A, Bargalló N, Sahuquillo J (2006) Hippocampal head atrophy after traumatic brain injury. Neuropsychologia 44(10):1956–1961. https://doi.org/10.1016/j.neuropsychologia.2005.11.007

Arora P, Singh K, Kumari M, Trivedi R (2022) Temporal profile of serum metabolites and inflammation following closed head injury in rats is associated with HPA axis hyperactivity. Metabolomics 18:28. https://doi.org/10.1007/S11306-022-01886-8

Atkins CM (2011) Decoding hippocampal signaling deficits after traumatic brain injury. Transl Stroke Res 2:546–555. https://doi.org/10.1007/S12975-011-0123-Z

Barzó P, Marmarou A, Fatouros P et al (1997) Contribution of vasogenic and cellular edema to traumatic brain swelling measured by diffusion-weighted imaging. J Neurosurg 87:900–907. https://doi.org/10.3171/JNS.1997.87.6.0900

Bennett ML, Viaene AN (2021) What are activated and reactive glia and what is their role in neurodegeneration? Neurobiol Dis 148:105172. https://doi.org/10.1016/J.NBD.2020.105172

Bramlett HM, Dietrich WD (2002) Quantitative structural changes in white and gray matter 1 year following traumatic brain injury in rats. Acta Neuropathol 103:607–614. https://doi.org/10.1007/S00401-001-0510-8

Budde MD, Janes L, Gold E et al (2011) The contribution of gliosis to diffusion tensor anisotropy and tractography following traumatic brain injury: validation in the rat using Fourier analysis of stained tissue sections. Brain 134:2248–2260. https://doi.org/10.1093/brain/awr161

Burda JE, Bernstein AM, Sofroniew MV (2016) Astrocyte roles in traumatic brain injury. Exp Neurol 275:305. https://doi.org/10.1016/j.expneurol.2015.03.020

Chu T, Zhang YP, Tian Z et al (2019) Dynamic response of microglia/macrophage polarization following demyelination in mice. J Neuroinflammation 16:1–22. https://doi.org/10.1186/S12974-019-1586-1

Clément T, Lee JB, Ichkova A et al (2020) Juvenile mild traumatic brain injury elicits distinct spatiotemporal astrocyte responses. Glia 68:528–542. https://doi.org/10.1002/glia.23736

Ferrara M, Bertozzi G, Volonnino G et al (2022) Glymphatic system a window on TBI pathophysiology: a systematic review. Int J Mol Sci 23:9138. https://doi.org/10.3390/ijms23169138

Giordano KR, Denman CR, Dubisch PS et al (2021) An update on the rod microglia variant in experimental and clinical brain injury and disease. Brain Commun 3:fcaa227. https://doi.org/10.1093/braincomms/fcaa227

Graham NSN, Sharp DJ (2019) Understanding neurodegeneration after traumatic brain injury: from mechanisms to clinical trials in dementia. J Neurol Neurosurg Psychiatry 90:1221–1233. https://doi.org/10.1136/jnnp-2017-317557

Haber M, Hutchinson EB, Sadeghi N et al (2017) Defining an analytic framework to evaluate quantitative MRI markers of traumatic axonal injury: preliminary results in a mouse closed head injury model. eNeuro 4:ENEURO.0164-17.2017. https://doi.org/10.1523/eneuro.0164-17.2017

Ho ML, Rojas R, Eisenberg RL (2012) Cerebral edema. AJR Am J Roentgenol 199:W258–W273. https://doi.org/10.2214/ajr.11.8081

Hulkower MB, Poliak DB, Rosenbaum SB et al (2013) A decade of DTI in traumatic brain injury: 10 years and 100 articles later. Am J Neuroradiol 34:2064–2074. https://doi.org/10.3174/ajnr.a3395

Johnson VE, Stewart W, Smith DH (2010) Traumatic brain injury and amyloid-β pathology: a link to Alzheimer's disease? Nat Rev Neurosci 11:361. https://doi.org/10.1038/nrn2808

Jorge RE, Acion L, Starkstein SE, Magnotta V (2007) Hippocampal volume and mood disorders after traumatic brain injury. Biol Psychiatry 62(4):332–338. https://doi.org/10.1016/j.biopsych.2006.07.024

Karperien AL, Jelinek HF (2015) Fractal, multifractal, and lacunarity analysis of microglia in tissue engineering. Front Bioeng Biotechnol 3:51. https://doi.org/10.3389/fbioe.2015.00051

Karve IP, Taylor JM, Crack PJ (2016) The contribution of astrocytes and microglia to traumatic brain injury. Br J Pharmacol 173:692–702. https://doi.org/10.1111/bph.13125

Kumar M, Arora P, Sandhir R (2022) Biphasic role of microglia in healthy and diseased brain. Biol Glial Cells Recent Adv 507–537. https://doi.org/10.1007/978-981-16-8313-8_18

Lee TH, Yang JT, Lin JR et al (2020) Protective effects of ischemic preconditioning against neuronal apoptosis and dendritic injury in the hippocampus are age-dependent. J Neurochem 155:430–447. https://doi.org/10.1111/jnc.15029

Liddelow SA, Guttenplan KA, Clarke LE et al (2017) Neurotoxic reactive astrocytes are induced by activated microglia. Nature 541:481. https://doi.org/10.1038/nature21029

Lowenstein DH, Thomas MJ, Smith DH, McIntosh TK (1992) Selective vulnerability of dentate hilar neurons following traumatic brain injury: a potential mechanistic link between head trauma and disorders of the hippocampus. J Neurosci 12:4846–4853. https://doi.org/10.1523/jneurosci.12-12-04846.1992

Mira RG, Lira M, Cerpa W (2021) Traumatic brain injury: mechanisms of glial response. Front Physiol 12:740939. https://doi.org/10.3389/fphys.2021.740939

Mohamed AZ, Corrigan F, Collins-Praino LE, Plummer SL, Soni N, Nasrallah FA (2020) Evaluating spatiotemporal microstructural alterations following diffuse traumatic brain injury. NeuroImage Clin 25:102136. https://doi.org/10.1016/j.nicl.2019.102136

Newcombe VFJ, Williams GB, Nortje J et al (2008) Concordant biology underlies discordant imaging findings: diffusivity behaves differently in grey and white matter post acute neurotrauma. Acta Neurochir Suppl 102:247–251. https://doi.org/10.1007/978-3-211-85578-2_47

Ng SY, Lee AYW (2019) Traumatic brain injuries: pathophysiology and potential therapeutic targets. Front Cell Neurosci 13:528. https://doi.org/10.3389/fncel.2019.00528

Pirici D, Mogoantă L, Mărgăritescu O, Pirici I, Tudorică V, Coconu M (2009) Fractal analysis of astrocytes in stroke and dementia. Rom J Morphol Embryol 50(3):381–390

Plata A, Lebedeva A, Denisov P et al (2018) Astrocytic atrophy following status epilepticus parallels reduced Ca2+ activity and impaired synaptic plasticity in the rat hippocampus. Front Mol Neurosci 11:215. https://doi.org/10.3389/fnmol.2018.00215/BIBTEX

Ren H, Lu H (2019) Dynamic features of brain edema in rat models of traumatic brain injury. NeuroReport 30:605. https://doi.org/10.1097/wnr.0000000000001213

Rodriguez-Paez AC, Brunschwig JP, Bramlett HM (2005) Light and electron microscopic assessment of progressive atrophy following moderate traumatic brain injury in the rat. Acta Neuropathol 109:603–616. https://doi.org/10.1007/S00401-005-1010-Z

Schiweck J, Eickholt BJ, Murk K (2018) Important shapeshifter: mechanisms allowing astrocytes to respond to the changing nervous system during development, injury and disease. Front Cell Neurosci 12:261. https://doi.org/10.3389/fncel.2018.00261

Siedler DG, Chuah MI, Kirkcaldie MTK et al (2014) Diffuse axonal injury in brain trauma: insights from alterations in neurofilaments. Front Cell Neurosci 8:429. https://doi.org/10.3389/fncel.2014.00429

Sierra A, Laitinen T, Gröhn O, Pitkänen A (2015) Diffusion tensor imaging of hippocampal network plasticity. Brain Struct Funct 220(2):781–801. https://doi.org/10.1007/S00429-013-0683-7

Su E, Bell M (2016) Diffuse axonal injury. Encycl Neurol Sci 998–999. https://doi.org/10.1016/B978-0-12-385157-4.00326-2

Taylor SE, Morganti-Kossmann C, Lifshitz J, Ziebell JM (2014) Rod microglia: a morphological definition. PLoS ONE 9:e97096. https://doi.org/10.1371/journal.pone.0097096

Trivedi R, Khan AR, Rana P et al (2012) Radiation-induced early changes in the brain and behavior: serial diffusion tensor imaging and behavioral evaluation after graded doses of radiation. J Neurosci Res 90:2009–2019. https://doi.org/10.1002/jnr.23073

Vespa PM, McArthur DL, Xu Y, Eliseo M, Etchepare M, Dinov I, Alger J, Glenn TP, Hovda D (2010) Nonconvulsive seizures after traumatic brain injury are associated with hippocampal atrophy. Neurology 75(9):792–798. https://doi.org/10.1212/wnl.0b013e3181f07334

Winkler EA, Minter D, Yue JK, Manley GT (2016) Cerebral edema in traumatic brain injury: pathophysiology and prospective therapeutic targets. Neurosurg Clin N Am 27:473–488. https://doi.org/10.1016/j.nec.2016.05.008

Witcher KG, Bray CE, Dziabis JE et al (2018) Traumatic brain injury-induced neuronal damage in the somatosensory cortex causes formation of rod-shaped microglia that promote astrogliosis and persistent neuroinflammation. Glia 66:2719. https://doi.org/10.1002/glia.23523

Young K, Morrison H (2018) Quantifying microglia morphology from photomicrographs of immunohistochemistry prepared tissue using ImageJ. JoVE 2018:e57648. https://doi.org/10.3791/57648

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

Palkin Arora[1,2] · Richa Trivedi[1] · Megha Kumari[1,3] · Kavita Singh[1] · Rajat Sandhir[2] · Maria M. D'Souza[1] · Poonam Rana[1]

✉ Richa Trivedi
richa.inmas@gov.in

[1] Radiological, Nuclear and Imaging Sciences (RNAIS), Institute of Nuclear Medicine and Allied Sciences (INMAS), DRDO, Delhi, India

[2] Department of Biochemistry, Panjab University, Chandigarh, India

[3] Department of Biotechnology, Delhi Technological University (DTU), Delhi, India

Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An Examination of A Single-Pass Heat Exchanger's Parametric Design Using Coding

Pardeep kumar Monu
Ganga Institute of Technology & Management (GITAM)
Naveen Kumar Naveen (naveenkumardce@gmail.com)
Delhi Technological University (DTU), India
Raj Kumar Raj
Deenbandhu Chhotu Ram University of Science and Technology (DCRUST), Sonipat, Haryana, India
Amit Sharma Amit
Deenbandhu Chhotu Ram University of Science and Technology (DCRUST), Sonipat, Haryana, India
Anju Anju
Kishan Lal Public (KLP) college, Rewari, Haryana, India

Method Article

Keywords: Hydrocracker unit heat exchanger; Design data book; Thermo-Hydraulic Structure; Coding; Forced convection

Posted Date: February 9th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3909292/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

The main objective of this real-life example is to help readers who may have just a basic understanding of engineering concepts and a limited amount of expertise with the exchange of heat equipment. The procedure of creating a structure, a thermal transfer area, as well as pressure decreases whereas determining whether the inferred variables satisfy every need is referred to as Thermo-Hydraulic Structure originality. Analyzing many design variables is necessary for the hydrocracker unit heat exchanger, commonly known as the compact thermal exchanger. The design phases and assumptions are covered in this research as well. Borlando C++ program is used for parametric research evaluations. The intricate connections between fluid mechanics and heat transfer are also shown in this case study, drawing from a variety of scholarly research papers and industry recognized design data book. The design seeks to produce a high rate of heat transfer rather than a large pressure drop. The coding is essential to the design of densely packed heat exchangers and represents a significant improvement over previous designs.

1. Introduction

This section describes computer programming on design of Backwash Cooler in turbo C++. In this section rating of Backwash Cooler by LMTD method in turbo C++ are presented and flow chart are also presented Kays W.M., London A.L [1,3]; Kuppan T[6].

1.1 Air-cooled heat exchanger

Heat release is a necessary part of any cycle that produces power or uses cooling. This also holds true for internal combustion engines, computers, electronic devices, and other cycles in chemical and process plants. The majority of the energy in fuel is discarded by the cooling system and emissions due to the effectiveness of current car engines. A fossil-fired power plant with an efficiency of roughly 40% needs to discard over 40% of the heat intake via the cooling system. In nuclear facilities with lower efficiency, even more heat needs to be discarded. In contemporary integrated cycle power plants, a notably reduced amount of heat is discarded.

The hydrosphere was once the most widely employed heat sink in manufacturing processes. Directing water from a river, dam, or ocean to a plant heat exchanger and then returning the warmed water to its origin was the easiest and least expensive way to cool a system. The allowed increase in temperature of this type of chilled water is frequently restricted in developed nations, which limits the use of natural water for cooling purposes. Air-cooled heat exchangers are used in industrial and chemical industries, automobiles, cooling systems, and chilling facilities where fluids with temperatures of at least 59°C need to be cooled. When cooling water is either highly costly or unavailable, the use of air-cooled or dry-cooling systems in industry or power stations makes sense.

Heat is transferred from the process fluid to the cooling airstream in a cooled by air. Heater, also known as an air cooler, using expanded surfaces or finned tubes. The atmospheric wet bulb temperature affects

the performance of wet-cooling systems, whereas the cooled by air thermal exchangers' output is based on the air's dry bulb temperature. The temperature of the dry bulb fluctuates more dramatically both daily and seasonally and is greater than that of the moist bulb. In many situations, dry/wet or wet/dry cooling arrangements provide the best choice. Small cooled air heat exchangers, also known as compact heat exchangers according to Kays [1, 3], find use in computers and other electronic equipment, vehicles (radiators, oil coolers, intercoolers), air conditioning, and freezing.

1.2 Backwash cooler

In many fields of engineering, heat transfer takes place between two fluids that are at different temperatures and divided by a solid wall. The phrase "heat converter" refers to the apparatus that facilitates this transfer; particular uses include chemical manufacturing, power generation, waste heat recovery, and space heating and cooling. Germany invented hydrocracking for the first time in 1915 in order to produce liquid fuels from their own coal reserves. In Leuna, Germany, the first hydrocracking factory that could be regarded as an industrial facility started operating in 1927. Similar initiatives to transform coal into liquid fuels were undertaken in France, Great Britain, and other nations.

Standard Oil of New Jersey and I.G. Farbenindustrie of Germany worked together between 1925 and 1930 to develop hydrocracking technology that could turn heavy petroleum liquids into fuels. These procedures were highly costly and needed temperatures above 375 °C and pressures between 200 and 300 bar.

British company Imperial Chemical Industries created a two-step hydrocracking method in 1939. In Germany, Great Britain, and the USA during World War II (1939–1945), two-stage hydrocracking methods were crucial to the production of aviation fuel.

Hydrocracking technology lost significance after World War II. The Middle East's huge crude oil supply eliminated the incentive to turn coal into liquid fuels. Hydrocracking was significantly less cost-effective than newly developed fluid catalytic cracking techniques when it came to turning high-boiling gasoline oils into fuels.

In the United States, hydrocracking expanded quickly in the late 1960s and early 1970s. The method of hydrocracking reached maturity by the mid-1970s, at which point its expansion started to slow down. After that, the United States' hydrocracking industry grew slowly. Nonetheless, hydrocracking had notable expansion throughout Europe, the Asia-Pacific area, and the Middle East concurrently.

In 2001, there were roughly 155 hydrocrackers in operation throughout the world, processing 4,000,000 barrels (550,000 metric tons) of feedstock daily. In 2009, the hydrocrackers in the United States had a daily feedstock treatment capacity of 1,740,000 barrels, or 238,000 metric tons.

1.2.1 Backwash cooler requirement

In the hydrocracker unit of an oil or petroleum refinery, the big hydrocarbons must be cracked into smaller bits by a backwash cooler or circular tube with a circular fin. Three main chemical reactions are involved

in catalytic hydrocracking:

- Creating lower-boiling, lower molecular weight hydrocarbons through the cracking of high-boiling, high molecular weight hydrocarbons present in petroleum crude oil.
- Using hydrogenation to produce saturated hydrocarbons, also known as paraffins or alkanes, from unsaturated hydrocarbons (whether from the initial feedstock or created during the cracking of the high-boiling, high molecular weight feedstock hydrocarbons).
- Hydrogenating any sulfur, nitrogen, or oxygen compounds present in the initial feedstock to produce gaseous ammonia, water, and hydrogen sulfide.

There are too many intricate reactions involved in the aforementioned basic processes to go into depth about each one. Examples of those complex reactions include the following four reactions:

- Reaction 1: Hydrogenated rings are created when aromatic compounds are added to hydrogen. Acid catalysts are then easily used to break these.
- Reaction 2: Double bonds are formed, bigger paraffins are broken up into smaller pieces, and paraffinic rings are opened by acid catalyst cracking.
- Reaction 3: To create paraffins, hydrogen is added to olefinic double bonds.
- Reaction 4: Isomerization of straight- and branched-chain paraffins is the fourth reaction.

Hydrocracking catalysts have two functions, namely hydrogenation and cracking, and are made of active metals on solid, acidic supports. The metals offer the hydrogenation function, while the acid catalyst support provides the cracking function.

1.2.2 Utilizing a backwash cooler

In the hydrocracker unit of an oil refinery, backwash coolers are employed. This air-cooled heat exchanger type is known as ACHE.

Hydrocracking became profitable in the early 1960s for several reasons:

- The auto industry started producing cars with improved performance that needed gasoline with a higher octane rating.
- The process of fluid catalytic cracking grew quickly in response to the need for high-octane gasoline. But in addition to producing gasoline, fluid catalytic cracking also yields cycle oil, a high-boiling byproduct that is extremely challenging to recycle for additional cracking. But that cycle oil can be cracked by hydrocracking.
- The 1950s saw a rise in the demand for diesel and jet fuel due to the conversion of railroad steam engines to diesel engines and the arrival of commercial jet aircraft. The ability of hydrocracking to yield diesel oil, jet fuel, or gasoline made hydrocrackery a desirable addition to petroleum refineries.

- Compared to the older catalysts, zeolite-based catalysts, which were developed and brought to market between 1964 and 1966, performed substantially better. Above all, they made it possible to operate at lower pressures than were previously achievable with catalysts. The improved catalysts allowed for higher performance and lower operating pressures, which led to far more cost-effective hydrocrackers.

Air-cooled heat exchangers are used in industrial and chemical plants, automobiles, air conditioning systems, and refrigeration facilities where fluids with temperatures of at least 60^0C need to be cooled. When cooling water is either extremely expensive or unavailable, the use of air-cooled or dry-cooling systems in industry or power plants makes sense.

Heat exchange occurs between two fluids separated by a solid wall and at different temperatures in many engineering applications. The equipment that makes this exchange possible is referred to as a "heat exchanger"; particular applications include power generating, waste heat recovery, chemical processing, and space heating and cooling. Heat transfer is a problem that many industries face, and because there are so many different application areas, there are many variations in the details. Since the calculation concepts that underlie the difficulty of building a heat exchanger are universal, the purpose of this study is to provide design calculations rather than to address the peculiarities of individual situations and unique cases [2].

A Borlando C++ computer software for a small heat exchanger circular tube with circular fin design was written. Heat exchanger design calculations typically appear to be complicated, requiring repetitive computations using fixed or assumed data. Since the design process involves a lot of thermodynamic and fluid dynamic characteristics, employing computer software will satisfy this requirement and preserve time.

Plant managers can use the software to design compact heat exchangers without the need for intricate design calculations, as well as make necessary modifications to the current heat exchanger, such as blocking blown-out tubes without impairing operation or changing fluid flow rate in accordance with other processes.

The goal of this article is to help practicing engineers categorize the numerous intricate questions that come up during design calculations and to help them apply their formal training in fluid movement and heat transfer to real-world issues that arise during the design, selection, testing, or installation of lightweight heat exchangers.

Heat transfer occurs between two fluids separated by a solid wall and at different temperatures in many engineering applications. The equipment that makes this exchange possible is referred to as a "heat exchanger"; particular applications include power generating, waste heat recovery, chemical processing, and space heating and cooling.

The issue of heat transfer arises in many different sectors, and due to the wide range of applications, there are innumerable variations in the specifics.

In an oil refinery, the hydrocracking unit uses a backwash chiller to break up the hydrocarbons into smaller pieces. There exist multiple varieties of cracking:

• Thermo breaking: combustion, the gas steam, and visbreaking.

• Chemical breaking: hydrocracking and fluid chemical cracking.

Input (Larger HC)	Output (Smaller HC)
Heavy oil	Gasoline, Kerosene oil

1.3 Design of Tube Fin Heat Exchanger

The designs of Hesselgreaves [7] and Muller [15] provide excellent descriptions of air-cooled heat exchangers and compact heat exchangers. Air-cooled heat exchangers (ACHEs) cost two to three times more than water-cooled heat exchangers for the same task, based on bare-tube heat transfer area. This is primarily due to two factors. First, air has a far lower heat transfer coefficient than water due to its significantly poorer thermal conductivity. Second, the mean temperature difference (MTD) is always less for an ACHE because ambient temperatures are always greater than design water temperatures, particularly at relatively low process-fluid outlet temperatures. The heat transfer area of an ACHE is significantly greater than that of a water-cooled heat exchanger for the same duty because of these two characteristics. The bigger size also necessitates a complex structural support system, which drives up the cost. Compared to air cooling, the operational costs of water cooling are substantially higher. These consist of the initial raw water's purchase price, makeup water and treatment chemicals, the plant cooling tower's allocated cost, and the pumping cost. The amount of power needed to create air flow across the tube bundles is all that constitutes the operational cost for ACHEs. A more advanced model has been created by Xie, Wang, and Sunden [12] to forecast the features of the j and f factors for tube-fin heat exchangers. The economics of air cooling are further favored by the rising running costs of water-cooled heat exchangers when water becomes more limited. Khan [11] presented a clear analytical method for calculating the convective heat transfer coefficient.

- Both the LMTD approach and the ε-NTU method were utilized in the heat exchanger design; figure 3.1 illustrates this process.
- Hudson is the top manufacturer satisfying ACHE's worldwide demand.

1.4 Objective

The aforementioned study's influences are as follows:

1. There are no available design studies for air-cooled heat exchangers (ACHEs).

2. There aren't many vendors available.
3. The design of the ACHE/Backwash cooler is elegant.
4. If water is not available (for example, in a dry environment), an air-cooled heat exchanger may be utilized.
5. The ACHE/Backwash cooler's design data is intellectual in nature.

The work's objectives are:

1. Backwash cooler design for hydrocracking unit catalytic cracking.
2. Effects of pipe length, pitch, fin density, diameter, and elevation were studied parametrically.

This work involves the design analysis of a backwash cooler, which is used to cool hydrocarbons like oil. Each backwash cooler has a single circular fin, and the backwash cooler is examined and coded.

2. Presumption

A number of elements that affect the tube side heat transfer coefficient, which in turn affects the tube side thermal exchange rate, must be taken into account by the engineer designing it.

1. With fins, air moves from bottom to highest with the aid of a forced draft fan, increasing the heat transfer coefficient compared to natural convection with other fins [4].
2. The coefficients of heat transfer and turbulence are influenced by tube layout patterns; for example, equilateral triangle pitch produces more turbulence than square pitch. Additionally, the heat transfer coefficient for triangular pitch is around 23% higher than for square pitch under same flow and tube size conditions [7].
3. Heat transfer will be at its highest as fin density rises because more air will come into contact with the tube more often [4].
4. Additionally, tube dimensions, clearance, and fluid-flow properties affect the air end coefficient [3].
5. The air side flow area varies over the bundle diameter depending on the fin concentration, tube diameter, and length of pipe [7].
6. Certain words are frequently used interchangeably while discussing heat exchanger specification challenges and their solutions. "Evaluation," "layout," and "choice" are among them. The term "Evaluation" refers to the mathematical process that takes as inputs the fluid characteristics, the heat exchange mechanism parameters, the entry flow rates and temperatures, and the outlet temperatures and heating duty (if the exchanger length is specified) or the necessary heat exchanger dimensions [13].

"Layout" is the process of determining all the constructional parameters required for an exchanger to meet particular heat duty requirements and air side and tube side pressure drop constraints. "Choice" describes the process of selecting a heat exchanger from a range of exchangers that are already in

place. Most engineers "cut and try" when using empirical data, depending instead on their experience and intuition to infer new designs from tested units. For the less experienced designer, however, a technique for analysis is simpler to follow because it illustrates the fundamental links. In this work, we will primarily focus on thermal evaluation using the analytical methodology proposed by W.M. Kays, A.L. London [4], and Wolverine [12]. Using the Borlando C++ application, all backwash cooler design computations are completed. The majority of thermal exchangers were designed under initial circumstances where the following characteristics were known or presumed to exist:

a. The fluids' velocity
b. The variation of fluid temperatures
c. Length of pipe and pipe configuration

It must create a layout for the ideal exchanger based on the details provided above in order to achieve the necessary process requirements. Typically, one needs to ascertain the subsequent outcomes.

a. Quantity of tubes
b. The pace of heat transmission
c. The factor of total flow of heat
d. Decrease in pressure on the outside and pipe sides

Three primary stages comprised the layout computations.

a. Determining the outermost geometry
b. Calculating and verifying the heating output
c. Calculating and verifying the tube and air side pressure reduction.

TABLE 1

INPUT DATA FOR BACKWASH COOLER

	Tube side	Air side
Fluid	Crude oil(Hydrocarbon, HC)	Air
Inlet temperature(^{0}C)	162	45
Outlet temperature(^{0}C)	80	59.6
Flow rate(m^3/hr)	13.16	161.11
Heat load(kW)	330.404	330.404
Pressure drop(kPa)	0.4(inlet) & 2.5(0utlet)	11.22
Max working pressure(kPa)	7.5	101.325

TABLE II

INDUSTRIAL DATA

Result	Panipat refinery data(IOCL)
External diameter of the pipe, d_o	25mm
Internal diameter of the pipe, d_i	20mm
Tube length, m	10.5
Total no. of tubes	276=46×6
Number of fin/m	433
Tube pitch, p_t	60mm (0.060m)
Fin Thickness, t_f	0.35mm (0.00035m)
Fin Height, l_f	12.5mm (0.0125m)
Root fin diameter, d_r	50mm (0.050m)
Heat Load,, kW	330.404
Pipe side pressure difference ,kPa	0.3808
Decrease in pressure on the air side ,kPa	11.22

3. Backwash cooler rating algorithm using a flow chart

The rating problem's fundamental steps, which include determining the following parameters, are depicted in figure 3.1. All these design process and flow chart are taken from several design books Kays W.M., London A.L [1,3]; Kuppan T[6]; Hesselgreaves John E.[7], Shah Ramesh K., Sekulic Dusan P.[8]; Professor John R.Thom [9].

1. Geometrical features of the surface. These consist of the following::
   A_o, $A = A_p + A_f$, D_h, σ, $A_{fr,airside}$, $A_{fr,tubeside}$, σ_1, σ_2
   Specialized dimensions employed for correlations of pressure reduction and circulation of heat are also provided.
2. Find each fluid's thermophysical characteristics at its bulk mean temperature, including the qualities of hot and cold fluids. The properties μ, cp, k, and Pr are required for solving the evaluation problem.
3. By determining the Reynolds number and any other pertinent dimensionless groups, find the j or Nu as well as f of the heat exchange surfaces on either side of an exchanger.
4. Estimate the f, j, and Nu factors.
5. Using Nu or j, determine the heat exchange factor for each of these two fluid flows.
6. Find the total surface efficiency (ηo) and fin efficiency (ηf).
7. Compute the wall thermal resistance, or Rw. Recognizing the fouling obstacles, Rf,h and Rf,c, on the hot and cold stream sides, respectively.
8. Determine the total heat conductance (UA).Obtain the heat duty.
9. Determine the drop in core pressure. For the varying fluid properties, the friction factor f on either side is rectified.

4. The backwash cooler's design program code

/*Rating of Backwash Cooler by LMTD method*/

```
#include<stdio.h>

#include<math.h>

#include<conio.h>

void main()

{
```

/*Surface Geometric Calculation*/

```
float pi,d_r,d_f,L_1,L_2,L_3,t_f,N_f,N_t,Z,X_t,X_l,d_i,V,X,Y;

float A_p,A_f,A,Ao_air,Ao_Tube,Afr_air,Afr_hc,alpha,Dh_air,sigma1_air,sigma2_hc;
```

pi=3.14,d_r=0.025,d_f=0.050,L_1=10.5,L_2=0.360,L_3=2.96,t_f=0.00035,

N_f=433,N_t=276,d_i=0.020,X_t=0.060,X_l=0.060,V=11.1888;

/*printf("\npi= ");

scanf("%f",&pi);

printf("\nDiameter (d_o),d_r= ");

scanf("%f",&d_r);

printf("\nRoot fin diameter,d_f= ");

scanf("%f",&d_f);

printf("\nTube inside diameter,d_i= ");

scanf("%f",&d_i);

printf("\nTube length,L_1= ");

scanf("%f",&L_1);

printf("\nflow length of air(height),L_2= ");

scanf("%f",&L_2);

printf("\nBreadth of heat exchanger,L_3= ");

scanf("%f",&L_3);

printf("\nfin thickness,t_f= ");

scanf("%f",&t_f);

printf("\nNo. of fin,N_f= ");

scanf("%f",&N_f);

printf("\nTotal no. of tubes,N_t= ");

scanf("%f",&N_t);

printf("\nZ= ");

```
scanf("%f",&Z);
printf("\nTransverse pitch,Xt= ");
scanf("%f",&Xt);
printf("\nLongitudnal pitch,Xl= ");
scanf("%f",&Xl);
   printf("\nVolume of heat exchanger,V= ");
scanf("%f",&V);*/
Ap=pi*dr*(L1-tf*Nf*L1)*Nt+2*(L2*L3-pi/4*dr*dr*Nt);
printf("\n\n\t Primary Area,Ap= %0.4f",Ap);
Af=(2*pi*(df*df-dr*dr)/4+pi*df*tf)*Nt*Nf*L1;
printf("\n\n\t Secondary Area,Af= %0.4f",Af);
A=Ap+Af;
printf("\n\n\t Total Heat tranfer surface area,A= %0.4f",A);
X=(Xt-dr)-(df-dr)*tf*Nf;
Y=2*((pow((pow(Xt/2,2)+pow(Xl,2)),0.5))-df-(df-dr)*tf*Nf);
if(X>Y)
{
Z=X;
Printf ("\n\n\t\t Z= %0.4f", Z);
}
else
{
Z=Y;
Printf ("\n\n\t\t Z= %0.4f", Z);
```

```
}
Ao_air=((L3/Xt-1)*Z+(Xt-dr)-(df-dr)*tf*Nf)*L1;
printf("\n\n\t Minimum free flow area Tube outside,Ao_air= %0.4f",Ao_air);
Ao_Tube=(pi/4)*di*di*Nt;
printf("\n\n\t Minimum free flow area tube inside,Ao_Tube= %0.4f",Ao_Tube);
Dh_air=(4*Ao_air*L2)/A;
printf("\n\n\t Hydraulic diameter,Dh_air= %0.4f",Dh_air);
alpha=A/V;
printf("\n\n\t Surface area density,alpha= %0.4f",alpha);
Afr_air=L1*L3;
Afr_hc=L2*L3;
printf("\n\n\t Frontal area,air side= %0.4f",Afr_air);
printf("\n\n\t Frontal area, tube side=%0.4f",Afr_hc);
sigma1_air=Ao_air/Afr_air;
sigma2_hc=Ao_Tube/Afr_hc;
printf("\n\n\t(sigma1)air= %0.4f",sigma1_air);
printf("\n\n\t(sigma2)tube= %0.4f",sigma2_hc);
/*Heat Transfer Calculation*/
/*Properties for Air at 52.5 0c(325.5k)*/
float rho_i_air,cp_air,mu_air,Pr_air;
/*Properties for Hydrocarbon at 121 0c(394k)*/
float rho_hc,mu_hc,k_hc,Pr_hc;
/*Thermal conductivity for fin (Al)*/
float k_fin;
```

```
/*Thermal conductivity for tube(carbon steel)*/
float k_tube;
/*fouling factor*/
float Rf_air,Rf_hc;
/*for air side calculation*/
float m_air,l_f,s,d_o,X_d;
float G_air,Re_air,j_air,h_o_air,f_air;
float r_fin,le_fin,m_fin,n_fin,phi_fin,eta_f_fin,eta_o_fin;
/*Tube side calculation*/
float G_hc, Re_hc, h_i_hc,m_hc;/*assume mu=mu_w*/
/* total coefficient of heat transfer */
float U_o, r_o, r_i, A_o, A_i;
/*Heat tranfer rate by LMTD*/
float Q,theta_m,theta_1,theta_2;
rho_i_air=1.0782,cp_air=1008,mu_air=0.00001964,Pr_air=0.7035;
rho_hc=804.3,mu_hc=0.002404,k_hc=0.109,Pr_hc=51.2471;
k_fin=225,k_tube=35,Rf_air=0.0004,Rf_hc=0.0006,m_air=161.11;
lf=0.0125,s=0.00196,Do=0.025,Pd=0.08485,r_fin=2,m_hc=13.16;
r_o=0.0125,r_i=0.010,A_o=0.8245,A_i=0.6597,theta_1=102.4,theta_2=35;
/*printf("\n\n\tDensity of air,rho_i_air = ");
scanf("%f",&rho_i_air);
printf("\n\n\tSpceific heat of air,cp_air = ");
scanf("%f",&cp_air);
printf("\n\n\tDynamic viscosity,mu_air = ");
```

```
scanf("%f",&mu_air);
printf("\n\n\tPrandtl Number of air,Pr_air = ");
scanf("%f",&Pr_air);
printf("\n\n\tDensity of hydrocarbon,rho_hc = ");
scanf("%f",&rho_hc);
printf("\n\n\tDynamic viscosity,mu_hc = ");
scanf("%f",&mu_hc);
printf("\n\n\tThermal conductivity of hydrocarbon,k_hc = ");
scanf("%f",&k_hc);
printf("\n\n\tPrandtl Number of hydrocarbon,Pr_hc = ");
scanf("%f",&Pr_hc);
printf("\n\n\tThermal conductivity of fin,k_fin = ");
scanf("%f",&k_fin);
printf("\n\n\tThermal conductivity of Tube,k_tube = ");
scanf("%f",&k_tube);
printf("\n\n\tFouling factor air side,Rf_air = ");
scanf("%f",&Rf_air);
printf("\n\n\tFouling factor hydrocarbon side,Rf_hc = ");
scanf("%f",&Rf_hc);
printf("\n\n\tmass flow rate of air,m_air = ");
scanf("%f",&m_air);
printf("\n\n\tLength of fin,l_f = ");
scanf("%f",&l_f);
printf("\n\n\tfin spacing,s= ");
```

```
scanf      ("%f",&s);
printf("\n\n\tTube outside diameter,d_o = ");
scanf("%f",&d_o);
printf("\n\n\tDiagonal pitch, X_d = ");
scanf("%f",&X_d);
printf("\n\n\tMass flow rate of hydrocarbon,m_hc = ");
scanf("%f",&m_hc);
printf("\n\n\t r_fin= ");
scanf("%f",&r_fin);
printf("\n\n\tTube inside radius,r_i = ");
scanf("%f",&r_i);
printf("\n\n\tTube outside radius,r_o = ");
scanf("%f",&r_o);
printf("\n\n\tTube inside surface area,A_i = ");
scanf("%f",&A_i);
printf("\n\n\tTube outside surface area,A_o = ");
scanf("%f",&A_o);
printf("\n\n\ttheta_1 = ");
scanf("%f",&theta_1);
printf("\n\n\ttheta_2 = ");
scanf("%f",&theta_2);*/
/*Air side calculation*/
G_air=m_air/(Afr_air*sigma1_air);
printf("\n\n\t Mass flow velocity of air,G_air= %0.4f",G_air);
```

```
Re_air=(G_air*Dh_air)/mu_air;
printf("\n\n\t Reynolds number for air,Re_air= %0.4f",Re_air);
j_air=0.134/((pow(Re_air,0.319))*(pow(lf/s,0.2))*(pow(tf/s,0.1134)));
printf("\n\n\t j= %0.4f",j_air);
h_o_air=(j_air*G_air*cp_air)/(pow(Pr_air,2/3.0));
printf("\n\n\t Convective heat transfer coefficient,h_o_air= %0.4f",h_o_air);
f_air=9.465/((pow(Re_air,0.316))*(pow(X_t/d_o,0.927))*(pow(X_d/X_t,0.515)));
printf("\n\n\t f= %0.4f",f_air);
/*fin Efficiency and Surface Efficiency calculation*/
le_fin=lf+(tf/2);
printf("\n\n\t le= %0.4f",le_fin);
m_fin=pow((2*h_o_air)/(k_fin*tf),1/2.0);
printf("\n\n\t m_fin= %0.4f",m_fin);
n_fin=exp(0.13*m_fin*le_fin-1.3863);
printf("\n\n\t n_fin= %0.4f",n_fin);
phi_fin=m_fin*le_fin*(pow(r_fin,n_fin));
printf("\n\n\t phi= %0.4f",phi_fin);
eta_f_fin=tanh(phi_fin)/phi_fin;
printf("\n\n\t fin efficiency,eta_f_fin= %0.4f",eta_f_fin);
eta_o_fin=1-((Af/A)*(1-eta_f_fin));
printf("\n\n\t overall efficiency,eta_o_fin= %0.4f",eta_o_fin);
/*Tube side calculation*/
G_hc=m_hc/(Afr_hc*sigma2_hc);
printf("\n\n\t Mass velocity of fluid,G_hc= %0.4f",G_hc);
```

```
Re_hc=(G_hc*d_i)/mu_hc;
printf("\n\n\t Reynolds number for hydrocarbon,Re_hc= %0.4f",Re_hc);
h_i_hc=(k_hc/d_i)*1.86*(pow((Re_hc*Pr_hc*d_i)/L_1,0.33));
printf("\n\n\t Convective heat transfer coeff. for hc,h_i_hc= %0.4f",h_i_hc);
/* total coefficient of heat transfer */
U_o =1 /((A_o/(eta_o_fin*h_i_hc*A_i))+((Rf_hc/eta_o_fin)*(A_o/A_i))
+((A_o/(2*pi*k_tube*L_1))*log(r_o/r_i))+(1/(eta_o_fin*h_o_air))
+(Rf_air/eta_o_fin));
printf("\n\n\t overall heat transfer coefficient, U_o= %0.4f",U_o);
/*Heat tranfer rate by LMTD method*/
theta_m = (theta_1-theta_2)/log(theta_1/theta_2);
printf("\n\n\t LMTD,theta_m= %0.4f",theta_m);
Q=(U_o*A_o*N_t*theta_m);
printf("\n\n\t Total heat transfer rate,Q= %0.4f",Q);
/*Pressure Drop Calculation*/
float gc,g,rh,rho_o_air,rho_m_air,v_hc;
float f,delta_p_air,delta_p_hc,hf;
gc=1,g=9.81,rh=0.0012,rho_o_air=1.05304,rho_m_air=1.076722
,v_hc=0.18869;
/*printf("\ng_c= ");
scanf("%f",&g_c);
printf("\nAcceleration due to gravity,g= ");
scanf("%f",&g);
printf("\nrh= ");
```

```
scanf("%f",&rh);
printf("\nrho_o_air= ");
scanf("%f",&rho_o_air);
printf("\nrho_m_air= ");
scanf("%f",&rho_m_air);
printf("\nv_hc= ");
scanf("%f",&v_hc);*/
/*Air side Pressure drop calculation*/
L2/rh;
printf("\n\n\t L2/rh= %0.4f",L2/rh);
delta_p_air=(G_air*G_air/2*gc*rho_i_air)*(f*L2/rh*rho_i_air*(1/rho_m_air)
+2*(rho_i_air/rho_o_air-1));
printf("\n\n\t Pressure drop air side,delta_p_air= %0.4f",delta_p_air);
/*Tube side Pressure drop calculation*/
f=16/Re_hc;
printf("\n\n\t Coefficient of friction, f=%0.4f",f);
hf=(4*f*L1*v_hc*v_hc)/(di*2*g);
printf("\n\n\t Head loss due to friction, hf= %0.4f",hf);
delta_p_hc=(rho_hc*g*hf);
printf("\n\n\t Pressure drop tube side,delta_p_hc= %0.4f",delta_p_hc);
getch();
}
```

5. Typical evaluations of outcomes of design

The following typical result has been presented based on the program code provided in the preceding section and the input parameter provided by the IOCL Panipat refinery.

Table 5.1.Typical backwash cooler result

Volume of heat exchanger, V (m^3)	= 9.0348
Principal surface area, A_p(m^2)	= 194.47
Area beneath the primary surface, A_f(m^2)	= 3762.87
Overall surface area for heat exchange, A(m^2)	= 3957.34
Density of surface area, σ(m^2/m^3)	= 438.01
Pitch diagonal, X_t(m)	= 0.060
Pitch longitudinally, X_l(m)	= 0.052
Air side lowest free flow area, $A_{o,airside}$(m^2)	= 6.63
Tube side lowest free flow area, $A_{o,tubeside}$(m^2)	= 0.0867
Hydraulic diameter, air side, $D_{h,airside}$(m)	= 0.0019
Hydraulic diameter, tube side, $D_{h,tubeside}$(m)	= 0.002
Frontal area, air side, $A_{fr,airside}$	= 31.08
Frontal area, tube side, $A_{fr,tubeside}$	= 0.86
$\sigma_{1,airside}$	= 0.21
$\sigma_{1,tubeside}$	= 0.1007

Table 5.2.Heat transfer result

Air side result	
Fluid mass velocity, G(kg/m^2×s)	= 21.78
Reynolds number, Re	= 2160
Component of friction, j	= 0.009711
Convective coefficient of thermal transfer, h_o (W/m^2×K)	= 269.57
Tube side result	
Fluid mass velocity, G(kg/m^2×s)	= 151.85
Reynolds number, Re	= 1263.32
Convective coefficient of thermal transfer, h_o (W/m^2×K)	= 49.65
Fin efficiency and Surface efficiency	
Fin effectiveness, η_f	= 0.68 or 68%
Total efficiency of the outer layer, η_o	= 0.6958 or 69.58%
Total heat exchange efficiency	
U_o (W/m^2×K)	= 23.13
Duration of heat exchange	
Q(kW)	= 330.404
Effectiveness	
ε	= 0.09499

Table 5.3.Pressure drop result

Air side result	
Fanning factor, f	= 0.3216
Pressure drop, Δp_a(kPa)	= 11.22
Tube side result	
Fanning factor, f	= 0.0127
Head loss due to friction, h_f(m)	= 0.048
Pressure drop, Δp_a(kPa)	= 0.3808

6. LIMITATIONS AND FUTURE SCOPE

6.1 Limitations

- The initial cost of an ACHE is much more than that of a water-cooled unit since air has a much lower specific heat and heat transfer capacity than water.
- Comprehensive adaptation to winter plans must be implemented in colder regions to guard against below-freezing temperatures. The original investment expenses are further increased as a result.
- This technique of temperature can be as low as 3–5°C in areas cooled by water heaters, but an economical approach temperature between the outlet process fluid and ambient air is usually in the range of 10-15°C. When water is used for trimming cooling after air cooling, this disadvantage is mitigated. Because of its large heat-transfer area, an ACHE requires a larger plot area than a water-cooled exchanger. This disadvantage can be overcome by positioning an ACHE atop an exhaust rack so as not to waste any significant plot space.
- Because of the low specific heat of air, large volumes of air need to be forced across the tube bundles. This is accomplished by large-diameter air-circulation blades spinning rapidly and producing a lot of sound.
- Seasonal changes in air temperature have the potential to impact performance, necessitating the integration of costly control systems to guarantee steady operation.
- Large obstacles like buildings, should be avoided when placing ACHEs because air recirculation can occur and decrease effectiveness.
- The arrangement of ACHEs is fairly complicated. Consequently, there are fewer suppliers of ACHEs than there are for chilled water shells as well as tube heat exchangers.

6.2 Ideas for additional research

- Extensions in layout for multipass ACHE.
- The layout evaluation will make utilization of the multiphase fluid characteristics that have been determined.

7. CONCLUSIONS

Using Borland C++ programs, the layout theory for a backwash cooler—a circular exchange of heat having annular fins—has been established. Several inferences can be made from this work:

- C-language programs have been used in the construction of single pass backwash coolers.
- The fin-tube properties that affect thermal efficiency, such as fin elevation, fin separation, fin width, and the thermal conductance of the fin substance, are primarily responsible for the drop in pressure and the value of the coefficient of convective transfer of heat.
- Two crucial design factors for backwash coolers are tube pitch and tube length, which have an impact on the core's uniformity and flow characteristics.

References

1. Kays W.M., London A.L., “Compact Heat Exchangers”, McGraw-Hill Book Company,1955.
2. J.P. Holman, 1963, Heat Transfer,8th ed, McGraw-Hill Book Company.
3. Kays W.M., London A.L., “Compact Heat Exchangers”, McGraw-Hill Book Company, 1964.
4. Taborek J., Hewitt G.F., Afgan N., “Heat Exchangers”, McGraw-Hill Book Company, 1983.
5. Mukherjee R., “Effectively Design Air Cooled Heat Exchangers”, Chemical Engineering Progress (February 1997).
6. Kuppan T., “Heat Exchanger Design Handbook”, Marcel Dekker, Inc., 2000.
7. Hesselgreaves John E., “Compact Heat Exchangers”, Pergamon, Elsevier Science Ltd., 2001.
8. Shah Ramesh K., Sekulic Dusan P., “Fundamentals of Heat Exchanger Design”, John wiley & sons, Inc., 2003
9. Professor John R.Thom,© 2004, “Wolverine Tube Heat Transfer Data book -3”, Wolverine Tube Inc, www.Wolverine.com
10. Singh R N, George A K, Arafin S, “Specific heat ratio, Gruneisen parameter and Debye temperature of crude oil”, Journal of Physics D: Applied Physics 39 (2006) 1220-1225.
11. Khan W.A., Culham J.R., Yovanovichs M.M., “Convection heat transfer from tube banks in crossflow: Analytical approach”, International Journal of Heat and Mass Transfer 49 (2006) 4831-4838.
12. Xie Gongnan, Wang Qiuwang, Sunden Bengt, “Parametric study and multiple correlations on air-side heat transfer and friction characteristics of fin- and-tube heat exchangers with large number of large-diameter tube rows”, Applied Thermal Engineering 29,11 (2008).
13. Su Thet Mon Than, Khin Aung Lin, Mi Sandar Mon, “Heat Exchanger Design”World Academy of Science, Engineering and Technology 46, 2008.
14. Cengel Yunus A., “Heat Transfer”, McGraw-Hill Book Company, 2nd Edition.
15. Mueller A.C., “Air Cooled Heat Exchanger”, Rt. 1,Box 53 J, Church Hill, Maryland 21623,USA.

Figures

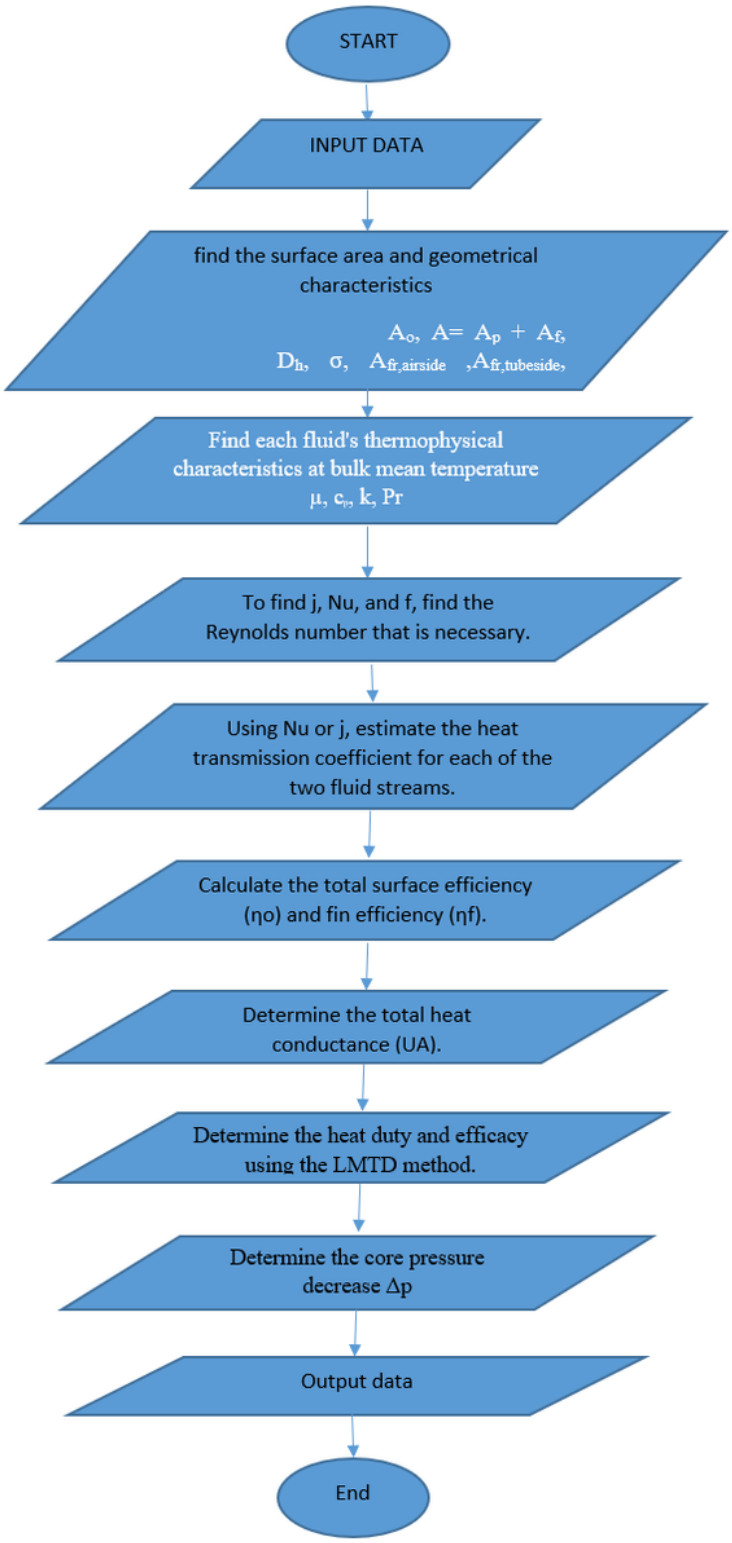


Figure 1

Figure 3.1 Diagrammatic flow chart of the backwash cooler rating

DOI: https://doi.org/10.52756/ijerr.2023.v36.028 *Int. J. Exp. Res. Rev.*, Vol. 36: 311-318 (2023)

Original Article | *Peer Reviewed* | *Open Access*

International Journal of Experimental Research and Review (IJERR)
© Copyright by International Academic Publishing House (IAPH)
ISSN: 2455-4855 (Online)
www.iaph.in

An Integrated Entropy-TOPSIS Approach for Electric Vehicle Selection

Kiran Pal[1]*, Dhruv Saraswat[1] and Neeraj Budhraja[2]

[1]Department of Mathematics, DITE Okhla Campus II, Delhi Skill & Entrepreneurship, Delhi, India; [2]Department of Mechanical Engineering, Delhi Technological University, Delhi, India

E-mail/Orcid Id:

KP, kiranpaldite@gmail.com; https://orcid.org/0000-0003-2743-3829,
DS, dhruvsaraswat10@gmail.com https://orcid.org/0000-0002-9349-4892
NB, neeraj_budhraja@yahoo.com, https://orcid.org/0000-0002-9349-4892

Article History:

Received: 13th Oct., 2023
Accepted: 16th Dec., 2023
Published: 30th Dec., 2023

Keywords:
Decision making, electric vehicle, entropy measure, MCDM, order of preference, TOPSIS

How to cite this Article:
Kiran Pal, Dhruv Saraswat and Neeraj Budhraja (2023). An Integrated Entropy-TOPSIS Approach for Electric Vehicle Selection. *International Journal of Experimental Research and Review*, *36*, 311-318.
DOI:
https://doi.org/10.52756/ijerr.2023.v36.028

Abstract: Electric vehicles (EVs) have gained a significant role in reducing emissions of harmful greenhouse gases, which can result in serious effects like global warming, deteriorating air quality, etc. Thus, it helps tackle concerns related to the environment and health of the general public. Many nations are switching towards this new era of electric mobility, thus promoting and contributing to the goal of sustainable development. Due to the rapid evolution and expanding scope of the electric vehicle market, selecting the best alternative between various existing EV models was tedious. In this paper, the electric vehicle in the Indian market context was considered along with the TOPSIS approach as a multi-criteria decision-making tool (MCDM) for choosing and ranking the best available choice for the electric vehicle present in the Indian market for the customer. The entropy method is used to obtain the weights associated with the criteria. For this study, thirteen electric vehicles have been selected as alternatives. This study contributes by giving a real preference order, considering a broad set of selection criteria, and evaluating the best alternative from the existing EVs.

Introduction

Increasing concern over environmental degradation, focus on sustainability, and detrimental health effects in recent years are motivating humans to switch from combustion vehicles to other alternatives to tackle this menace (Knez et al., 2019). Unfortunately, the vast majority of vehicles, even today, are vehicles that run on fossil fuels and release toxic substances into the environment, causing environmental degradation. In recent decades, the world has been under the threat of global warming and climate change as never before (Das et al., 2019; Ecer, 2021) due to non-renewable-based technologies such as internal combustion engine vehicles (Khan et al., 2020) Kumar and Alok (Kumar and Alok, 2020) claimed that the transport sector is the main actor in air pollution and ozone depletion ending in climate change by reason of greenhouse gas (GHG) emissions like CO, CH_4, N_2O, and CO_2; these have necessitated road transport electrification. Knowing the environmental concerns and technical advancements, combustion vehicles are being slowly replaced with cleaner ones, which use green energy alternatives that contribute to the overall betterment of the environment. But there is a ray of hope in this aspect. From the view of sustainable development, sustainable transport has several environmental, social, and economic advantages that could support sustainable development (Wei et al., 2020). Both to struggle with the environmental pollution problem and support sustainable development, many developed countries worldwide have shifted toward electric vehicles (EVs) (Babar et al., 2021). This is why EVs have been regarded as convenient solutions for improving hazardous gas reductions (Tran et al., 2020). Changing trends in automotive markets and ever-increasing environmental concerns have presented us with

*Corresponding Author: kiranpaldite@gmail.com

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

a whole new concept of electric vehicles. The recent century will be the century of electric vehicles. The onset of 21st century third decade has initiated a new era of electric vehicles (Secinaro et al., 2020). Electric cars have been increasingly accepted as a substitute for conventional cars.

Electric cars offer various advantages over conventional combustion vehicles, primarily characterized by their higher efficiency rates and minimal environmental impact (Zahoor et al., 2023). Amid the growing urban pollution and, consequently, the negative effects on people's health, governments across the globe are seeking effective solutions. This is why Electric vehicles are increasingly seen as an alternative to the conventional combustion vehicles.

Therefore, the widespread use of EVs will play a vital role in solving environmental and global economic problems in the automotive industry (Hoque et al., 2017). Notwithstanding, some authors noted that the environmental advantages of clean vehicles and the EU's zero-emission target on transportation could only be accomplished if the electricity is produced from clean resources like solar and wind (Neves et al., 2019). As per the report published by the International Energy Agency (IEA), the CO_2 equivalent emissions from EVs were about 38 million tons on a well-to-wheel basis in 2018, whereas the conventional fleets were about 78 million tons, respectively. Further, it is estimated that the demand for oil products will reduce by 4.3 million barrels per day thanks to EVs in the 2030s (International Energy Agency (IEA). Global EV Outlook 2019, 2020).

Today, the drawbacks of EVs, such as high cost, short battery range, and low top speed, have been partially eliminated. EVs have thus become one of the best alternatives for conventional versions (Ecer, 2021).

Consumers have recently adopted EVs (Samaie et al., 2020). Therefore, the number of EVs and electric mobility has exponentially increased, and this rise is still continuing (Emadi et al., 2005). In parallel with the e-car industry's rapid growth, the development of logistics services utilizing these vehicles has been observed (Yan et al., 2023). Besides their fundamental role in goods and people transportation, such vehicles exhibit zero environmental impact – an increasingly pertinent feature in contemporary society (Yan et al., 2023). It has been noted that fossil fuel vehicles consume more fuel within the urban environment than open roads, thereby significantly contributing to urban air pollution (Liu et al., 2023; Russo et al., 2021). This is being addressed through the incorporation of e-cars into urban logistics, emerging as a critical segment within green logistics (Strale, 2019).

Research Methodology

TOPSIS method

TOPSIS is a numerical method for solving multi-criteria decision-making problems (Wei et al., 2020). The method is based on the assumption that the chosen alternative should have the shortest Euclidean distance from the optimal solution and the shortest Euclidean distance from the negative positive solution Optimal Solution Edge Imaginary solution in which all the values of the components coincide with the highest value in database showing a satisfactory solution. The negative optimal solution is related to the concept of a search solution in which all feature values are the minimum attribute values in the database. Thus, TOPSIS provides a solution that is theoretically closest to the optimal solution and theoretically farthest from the worst-case solution. The basic multi-criteria decision-making process of the TOPSIS method is described to select the best alternative among the available ones.

Assumptions

In this study thirteen electric vehicle Tata Tiago EV (Ev1), MG Comet EV (Ev2), Mahindra XUV400 (Ev3), Tata Nexon EV (Ev4), Hyundai Kona Electric (Ev5), BMW i4 (Ev6), Kia EV6 (Ev7), MG ZS EV (Ev8), BYD E6 (Ev9), BYD Atto3 (Ev10), Tata Nexon EV Prime (Ev11), Hyundai Ioniq 5 Long Range RWD (Ev12), Hyundai Ioniq 5 Long Range AWD (Ev13) were taken under consideration with the following essential criteria: Price, Acceleration, Battery Capacity, Maximum Power, Range, Charging Time which has been described as below:

1. Price (Cr1): Price of the Electric vehicle varied across different showrooms in India (Price in ₹ Lakhs).
2. Acceleration (Cr2): Acceleration of the vehicle as measured from 0-100 kmph speed achieved in seconds (Acceleration (0-100 km/h) in seconds).
3. Battery capacity (Cr3): Battery capacity refers to the power stored in a single charge (Battery capacity (kWh)).
4. Maximum power (Cr4): Maximum power as measured in bhp of electric vehicle (Max Power (kW)).
5. Range (Cr5): Maximum range of electric vehicle in a single full charge of the vehicle (Range (km)).
6. Charging time (Cr6): the time taken by the vehicle to charge to full capacity (Charging time (0-80%) hours).

DOI: https://doi.org/10.52756/ijerr.2023.v36.028

Among the criteria thus considered, Price (Cr1) and Charging time (Cr6) are Cost criteria, whereas Acceleration (Cr2), Battery capacity (Cr3), Maximum power (Cr4) and Range (Cr5) are benefit criteria.

Table 1 displays a general specification decision matrix with three sparks plug manufacturing businesses. Rows represent options, while columns represent attributes.

The data for the following table has been taken from

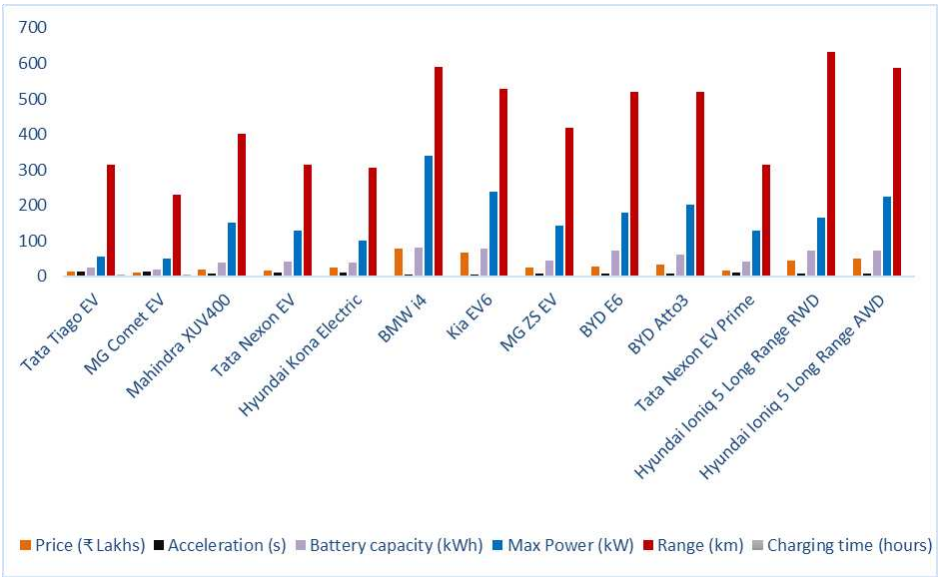


Figure 1. Comparison of various specifications of the thirteen electric vehicles.

website of car dekho (https://www.cardekho.com/).

Matrix rank: m*n

Results and Discussion

Steps Involved in Employing TOPSIS Methods

Step 1. Construction of the decision matrix

Create a decision matrix that comprises Alternative and Attributes for selecting the best alternative as represented in eq. (1),

Alternatives: i = 1,2,3,4,.....m

Attributes: j = 1, 2, 3,....n

Decision Matrix: X = [xij] (1)

Step 2. Calculate the normalized decision matrix

Elements of the matrix are divided by the square root of the sum of each squared element as depicted in eq. (2),

$$n_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} {x_{ij}}^2}} \quad (2)$$

Table 1. Decision matrix of various specifications of the thirteen electric vehicles.

Electric Vehicle/ Criteria	Cr1	Cr2	Cr3	Cr4	Cr5	Cr6
Ev1	12.04	11.9	24	55	315	5.5
Ev2	9.98	12.95	17.3	50	230	5.5
Ev3	18.99	8.3	39.4	150	400	0.833
Ev4	16.99	9.9	40.5	129	315	1
Ev5	24.04	9.7	39.2	100	305	1
Ev6	77.5	5.7	80.7	340	590	0.5166
Ev7	65.95	5.2	77.4	239	528	0.3
Ev8	23.38	8.5	44.5	143	419	0.833
Ev9	27.99	8.5	71.7	180	520	1
Ev10	33.99	7.3	60.48	201	521	1
Ev11	14.74	9.2	40.5	129	315	1
Ev12	44.95	7.6	72.6	165	631	0.3
Ev13	48.95	6.1	72.6	225	587	0.3

DOI: https://doi.org/10.52756/ijerr.2023.v36.028

Table 2. Normalized Decision Matrix

Electric Vehicle/Criteria	Cr1	Cr2	Cr3	Cr4	Cr5	Cr6
Ev1	0.0875	0.3750	0.1184	0.0854	0.1920	0.6698
Ev2	0.0725	0.4080	0.0854	0.0776	0.1402	0.6698
Ev3	0.1380	0.2615	0.1944	0.2328	0.2439	0.1014
Ev4	0.1234	0.3119	0.1998	0.2002	0.1920	0.1218
Ev5	0.1746	0.3056	0.1934	0.1552	0.1859	0.1218
Ev6	0.5630	0.1796	0.3982	0.5277	0.3597	0.0629
Ev7	0.4791	0.1638	0.3819	0.3709	0.3219	0.0365
Ev8	0.1699	0.2678	0.2196	0.2219	0.2554	0.1014
Ev9	0.2033	0.2678	0.3538	0.2794	0.3170	0.1218
Ev10	0.2469	0.2300	0.2984	0.3120	0.3176	0.1218
Ev11	0.1071	0.2899	0.1998	0.2002	0.1920	0.1218
Ev12	0.3266	0.2395	0.3582	0.2561	0.3847	0.0365
Ev13	0.3556	0.1922	0.3582	0.3492	0.3579	0.0365

The normalized decision matrix is acquired and shown in Table 2.

Step 4: Calculating the Criteria Weights

The Entropy Value is used by the Entropy technique to determine the weight values of the criterion. The Entropy Value and the criterion weight increase with increasing data dispersion. On the basis of different attributes, a weighted matrix was established and shown in Table 3. The steps in this method's evolution are as follows:

They are calculating the Entropy (E_i) Value as depicted in eq. (3). Here, all of the data in the normalized decision matrix have their natural logarithm values (ln) calculated. The normalized data is then multiplied by these values, and the resulting product is divided by the number of options' natural logarithm (ln(n)).

$$E_i = \frac{\sum_{j=1}^{n} p_{ij} \ln p_{ij}}{\ln n} \quad (3)$$

Calculate the weight values for the criteria. After completing the (1 − Ei) calculation, these values are totalled for each criterion. Next, the final weight of the criteria is determined as per eq. (4),

$$w_i = \frac{1-E_i}{\sum_{j=1}^{n}(1-E_i)} \quad (4)$$

Table 3. Weighted Matrix

Criteria	Cr1	Cr2	Cr3	Cr4	Cr5	Cr6
weight	0.149993	0.2217	0.198154	0.184525	0.216054	0.029574

Step 3. Weighted normalized decision matrix

The weighted normalized value V_{ij} was calculated as eq. (5),

$V_{ij}=w_{ij} * n_{ij}$, i=1,2,3,4……m, j=1,2,3,4….n (5)

where $V_{ij} = W_{ij\times}n_{ij}$

The obtained normalized weighted matrix is presented in Table 4.

Step 4. Determining the positive and negative ideal solutions

In the case of advantageous qualities (those with higher values that are desired for the specific application), V_j^+ represents the higher value of the attribute, as specified in eq. (6).

$$V_j^+ = \{(\sum_i^{max} v_{ij}/j\varepsilon J, \sum_i^{min} v_{ij}/j\varepsilon J')\ /i = 1,2,..N\} \quad (6)$$
$$= \{V_1^+, V_{2,}^+ V_{3\ldots\ldots}^+ V_M^+\}$$

In the case of non-beneficial attributes (i.e., those of which lower values are desired for the given application),V_j^- indicates the lowest value of the attribute as presented eq. (7),

$$V_j^- = \{(\sum_i^{min} v_{ij}/j\varepsilon J, \sum_i^{max} v_{ij}/j\varepsilon J')\ /i = 1,2,..N\} \quad (7)$$
$$= \{V_1^-, V_{2,}^- V_{3\ldots\ldots}^- V_M^-\}$$

where J = (j = 1, 2, …, M) /j is associated with beneficial attributes, and J’ = (j = 1, 2, …, M) /j is associated with non-beneficial attributes

The Positive and Negative ideal solutions are shown in Tables 5 and 6,

Step 5. Calculate the Difference measure (S_i^+ and S_i^-) for each alternative A_j from PIS and NIS

Difference measured given by Euclidean for each alternative have been calculated using eq. (8) and (9). The obtained results are tabulated in Table 7.

Table 4. Normalized Weighted Matrix (V_{ij})

Electric Vehicle/Criteria	Cr1	Cr2	Cr3	Cr4	Cr5	Cr6
Ev1	0.0131	0.0831	0.0235	0.0158	0.0415	0.0198
Ev2	0.0109	0.0905	0.0169	0.0143	0.0303	0.0198
Ev3	0.0207	0.0580	0.0385	0.0430	0.0527	0.0030
Ev4	0.0185	0.0692	0.0396	0.0369	0.0415	0.0036
Ev5	0.0262	0.0678	0.0383	0.0286	0.0402	0.0036
Ev6	0.0845	0.0398	0.0789	0.0974	0.0777	0.0019
Ev7	0.0719	0.0363	0.0757	0.0684	0.0695	0.0011
Ev8	0.0255	0.0594	0.0435	0.0410	0.0552	0.0030
Ev9	0.0305	0.0594	0.0701	0.0516	0.0685	0.0036
Ev10	0.0370	0.0510	0.0591	0.0576	0.0686	0.0036
Ev11	0.0161	0.0643	0.0396	0.0369	0.0415	0.0036
Ev12	0.0490	0.0531	0.0710	0.0473	0.0831	0.0011
Ev13	0.0533	0.0426	0.0710	0.0644	0.0773	0.0011

$$S_i^+ = d(V_{ij}, V_j^+) = \sqrt{(v_{ij} - v_j^+)^2} \quad (8)$$

and

$$S_i^- = d(V_i, V_j^-) = \sqrt{(v_{ij} - v_j^-)^2} \quad (9)$$

for i=1,2,3…….m

Table 5. Positive ideal solution (PIS)

V_j^+	0.0109	0.0905	0.0789	0.0974	0.0831	0.0011

Table 6. Negative ideal solution (NIS)

V_j^-	0.0845	0.0363	0.0169	0.0143	0.0303	0.0198

Step 6. Calculate the corresponding closeness coefficient (CC_i) of the alternatives

Table 7. Alternate from PIS and NIS.

Alternative	S_i^+	S_i^-
Ev1	0.1090	0.0863
Ev2	0.1178	0.0913
Ev3	0.0817	0.0813
Ev4	0.0863	0.0827
Ev5	0.0947	0.0735
Ev6	0.0895	0.1154
Ev7	0.0876	0.0918
Ev8	0.0800	0.0796
Ev9	0.0612	0.0968
Ev10	0.0666	0.0886
Ev11	0.0875	0.0829
Ev12	0.0736	0.0932
Ev13	0.0726	0.0949

The alternative's corresponding closeness coefficient (CCi) was calculated using eq. (10) as follows:

$$CC_i = \frac{S_i^-}{(S_i^+ + S_i^-)} \quad \text{for i=1,2,3…..m} \quad (10)$$

The obtained result is presented in Table 8.

Step 7. Rank the preference order or select the alternatives closest to 1

The TOPSIS approach was used to determine preference order, which is displayed in Table 9 and Figure 2.

The TOPSIS method has been successfully applied in the study. The determined findings have proven that BYD E6 had the uppermost priority (Rank 1), BYD Atto3 had the second uppermost (Rank 2), and Hyundai Ioniq 5 Long Range

AWD ranked third (Rank 3). In contrast, Hyundai Kona Electric (Rank 12) and MG Comet EV (Rank 13) rank lowest, taking into consideration the various criteria mentioned earlier (from a customer point of view).

DOI: https://doi.org/10.52756/ijerr.2023.v36.028

Table 8. Corresponding closeness coefficient (CC_i) of the alternatives.

Alternatives	CC_i	Alternatives	CC_i
Tata Tiago EV	0.4419	Hyundai Kona Electric	0.4370
MG Comet EV	0.4367	BMW i4	0.5633
Mahindra XUV400	0.4987	Kia EV6	0.5117
Tata Nexon EV	0.4894	MG ZS EV	0.4986
BYD E6	0.6127	BYD Atto3	0.5709
Tata Nexon EV Prime	0.4866	Hyundai Ioniq 5 Long Range RWD	0.5586
Hyundai Ioniq 5 Long Range AWD	0.5665		

Table 9. Obtained preference order.

Rank	Brand/Manufacturer	Rank	Brand/Manufacturer
1	BYD E6	**7**	Mahindra XUV400
2	BYD Atto3	**8**	MG ZS EV
3	Hyundai Ioniq 5 Long Range AWD	**9**	Tata Nexon EV
4	BMW i4	**10**	Tata Nexon EV Prime
5	Hyundai Ioniq 5 Long Range RWD	**11**	Tata Tiago EV
6	Kia EV6	**12**	Hyundai Kona Electric
		13	MG Comet EV

reflecting the multi-faceted nature of this decision-making

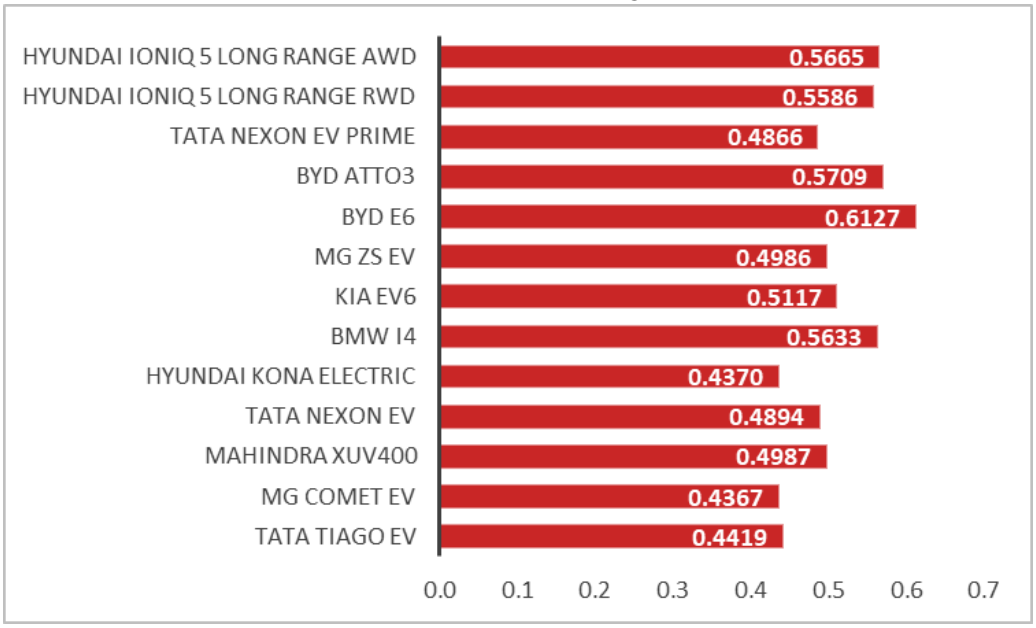


Figure 2. Corresponding closeness coefficient.

Conclusion

This paper on the selection of various electric vehicles present in the Indian market through the TOPSIS approach has provided valuable insights into the tedious process of selecting an electric vehicle that suits best from the consumer's point of view. Our paper encompassed a range of criteria such as cost, acceleration, battery capacity, charging time, etc., process.

Through applying the TOPSIS method of multi-criteria decision-making, we have systematically analyzed, assessed, and ranked the available electric vehicles in the market. This approach has allowed us to evaluate trade-offs (conflicting interests) objectively and make informed decisions.

In a nutshell, this study makes a good contribution to the ever-expanding and evolving market and knowledge

DOI: https://doi.org/10.52756/ijerr.2023.v36.028

space in the era of electric mobility. Also, it highlights how techniques related to multi-criteria decision-making can be used in this emerging field. The various insights gained from this paper will lead to more efficient choices among consumers and various other stakeholders of electric mobility.

Future scope

This study has also underlined the significance of employing TOPSIS (MCDM) in the context of electric vehicle selection and its potential for broader applications in decision-making across various areas. Focusing on the quantitative aspect of decision-making and comparing various criteria makes it possible to have a structured and data-driven approach for addressing multi-criteria decision problems.

Acknowledgements

The authors express their gratitude for all of the help they received from Delhi Technological University in Delhi, India.

Conflict of Interest

The authors declare no known conflict of interest for this article.

References

Babar, A. H. K., Ali, Y., & Khan, A. U. (2021). Moving toward green mobility: overview and analysis of electric vehicle selection, Pakistan a case in point. *Environment, Development and Sustainability*, *23,* 10994–11011. https://doi.org/10.1007/s10668-020-01101-5

Das, S., Bhattacharjee, A., Maity, M., Sahoo, R., & Mondal, J. (2019). Selection of Optimal Four Wheeler Applying AHP & TOPSIS Technique. *International Journal of Experimental Research and Review*, *20*, 10-27. https://doi.org/10.52756/ijerr.2019.v20.002

Ecer, F. (2021). Sustainability assessment of existing onshore wind plants in the context of triple bottom line: a best-worst method (BWM) based MCDM framework. *Environmental Science and Pollution Research*, *28,* 19677–19693. https://doi.org/10.1007/s11356-020-11940-4

Emadi, A., Rajashekara, K., Williamson, S. S., & Lukic, S. M. (2005). Topological overview of hybrid electric and fuel cell vehicular power system architectures and configurations. *IEEE Transactions on Vehicular Technology*, *54*(3), 763–770. https://doi.org/10.1109/TVT.2005.847445

Hoque, M. M., Hannan, M. A., Mohamed, A., & Ayob, A. (2017). Battery charge equalization controller in electric vehicle applications: a review. *Renewable and Sustainable Energy Reviews*, *75,* 1363–1385. https://doi.org/10.1016/j.rser.2016.11.126

International Energy Agency (IEA). Global EV Outlook 2019. Available at: https://www.iea.org/reports/global-ev-outlook-2019 [Accessed on 10 March 2020].

Khan, F., Ali, Y., & Khan, A. U. (2020). Sustainable hybrid electric vehicle selection in the context of a developing country. *Air Quality, Atmosphere & Health*, *13,* 489–499. https://doi.org/10.1007/s11869-020-00812-y

Knez, M., Zevnik, G. K., & Obrecht, M. (2019). A review of available chargers for electric vehicles: United States of America, European Union, and Asia. *Renewable and Sustainable Energy Reviews, 109,* 284–293. https://doi.org/10.1016/j.rser.2019.04.013

Kumar, R. R., & Alok, K. (2020). Adoption of electric vehicle: a literature review and prospects for sustainability. *Journal of Cleaner Production*, *253,* 119911. https://doi.org/10.1016/j.jclepro.2019.119911

Liu, J., Li J., Chen, Y., Lian, S., Zend, J., Geng, M., Zheng, S., Dong, Y., He, Y., Huang, P., Zhao, Z., Yan, X., Hu, Q., Wang, L., Yang, D., Zhu, Z., Sun, Y., Shang, W., Wang, D., Zhang, L., Hu, S., & Chen, X. (2023). Multi-scale urban passenger transportation CO_2 emission calculation platform for smart mobility management. *Applied Energy*, *331,* 120407, 2023. https://doi.org/10.1016/j.apenergy.202 2.120407

Neves, S. A., Marques, A. C., & Fuinhas, J. A. (2019). Technological progress and other factors behind the adoption of electric vehicles: empirical evidence for EU countries. *Research in Transportation Economics*, *74,* 28–39. https://doi.org/10.1016/j.retrec.2018.12.001

Russo, S. M., Voegl, J., & Hirsch, P. (2021). A multi-method approach to design urban logistics hubs for cooperative use. *Sustainable Cities and Society*, *69,* 102847. https://doi.org/10.1016/j.scs.2021.102847

Samaie, F., Meyar-Naimi, H., Javadi, S., & Feshki-Farahani, H. (2020). Comparison of sustainability models in development of electric vehicles in Tehran using fuzzy TOPSIS method. *Sustainable Cities and Society*, *53,* 101912. https://doi.org/10.1016/j.scs.2019.101912

Secinaro, S., Brescia, V., Calandr,a D., & Biancone, P. (2020). Employing bibliometric analysis to identify suitable business models for electric cars. *Journal of Cleaner Production*, *264,* 121503. https://doi.org/10.1016/j.jclepro.2020.121503

Strale, M. (2019). Sustainable urban logistics: What are we talking about? *Transportation Research Part A: Policy and Practice*, *130,* 745–751. https://doi.org/10.1016/j.tra.2019.10.002

Tran, D., Vafaeipour, M., Baghdadi, M., Barrero, R., Mierlo, J., & Hegazy, O. (2020) Thorough state-of-the-art analysis of electric and hybrid vehicle powertrains: topologies and integrated energy management strategies. *Renewable and Sustainable Energy Reviews*, *119,* 109596. https://doi.org/10.1016/j.rser.2019.109596

Wei, W., Cao, M., Jiang, Q., Ou, S. J., & Zou, H. (2020). What influences Chinese consumers' adoption of battery electric vehicles? A preliminary study based on factor Analysis. *Energies*, *13*(5), 1057. https://doi.org/10.3390/en13051057 https://www.cardekho.com [Accessed on 15 March 2022]

Yan, Z., Ding, H., & Chen, L. (2023). The analyzing the role of electric vehicles in urban logistics: A case of China. *Frontiers in Environmental Science*, *11,* 1128079. https://doi.org/10.3389/fenvs.2023.1128079

Zahoor, A., Mehr, F., Mao, G., Yu, Y., & Sápi, A. (2023). The carbon neutrality feasibility of worldwide and in China's transportation sector by e-car and renewable energy sources before 2060. *Journal of Energy Storage*, *61,* 106696. https://doi.org/10.1016/j.est.2023.106696

How to cite this Article:

Kiran Pal, Dhruv Saraswat and Neeraj Budhraja (2023). An Integrated Entropy-TOPSIS Approach for Electric Vehicle Selection. *International Journal of Experimental Research and Review*, *36*, 311-318.

DOI : https://doi.org/10.52756/ijerr.2023.v36.028

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Global Journal of Flexible Systems Management
https://doi.org/10.1007/s40171-024-00373-4

ORIGINAL RESEARCH

Analyzing Barriers in Adoption of Artificial Intelligence for Resilient Health Care Services to Society

Girish Kumar[1] · Rajesh Kumar Singh[2] · Vedpal Arya[3] · Shivam Kumar Mishra[1]

Received: 11 October 2023 / Accepted: 3 January 2024
© The Author(s) under exclusive licence to Global Institute of Flexible Systems Management 2024

Abstract *Artificial intelligence (AI) is emerging as an alternative solution in the healthcare sector, offering opportunities to enhance efficiency and optimize the utilization of precious resources. The nascent stage of AI application in the Indian healthcare sector has gained momentum during COVID-19, witnessing a surge in AI-based startups and companies specializing in diagnostic and prescriptive healthcare. This paper systematically analyzes barriers to AI application in the Indian context, incorporating the concept of flexibility, and proposes insights for fostering its adoption for resilient and sustainable healthcare practices. The identified barriers are drawn from literature and experts' inputs. These are further investigated using the decision-making trial and evaluation and laboratory methodology. This analysis not only categorizes barriers into cause-and-effect groups but also emphasizes the need for flexibility to adapt AI solutions to the dynamic healthcare sector. The paper underscores foundational barriers, including inadequate regulations, lack of awareness, high adaptation costs, and a scarcity of skilled AI expertise. In addition to managerial and social implications concerning regulation, implementation, economic viability, and data privacy, the study promotes flexibility as a key factor in addressing the evolving challenges in healthcare.*

✉ Rajesh Kumar Singh
rajesh.singh@mdi.ac.in

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

[2] Management Development Institute (MDI), Gurgaon, India

[3] Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore, Singapore

Keywords Artificial intelligence · Barriers · DEMATEL · Flexibility · Healthcare · Resilience · Sustainability

Introduction

Growing population and limited resources have posed a significant challenge to the healthcare sector across all countries. Process, people, production and technology, all these factors are exposed to a high level of external uncertainty. In health care sector, it is essential to deliver each and every product or service on accurate time as it is the question of life and death of patients (Vishwakarma et al., 2023). The healthcare supply chain (HSC) is a gigantic system of procedures and mechanisms that requires the well-timed production, dispersal, allocation of prescriptions and finally supply to patients (Abu Zwaida et al., 2021; Kumar et al., 2023). Developing flexibility across all processes is strategic requirement to survive in globally competitive market (De la Gala-Velásquez et al., 2023; Dhar et al., 2022; Sushil, 2015). In general, flexibility is the ability of a business to respond quickly and affordably to changes in the environment, technology, organization and strategy. As a result, it consists of initiatives aimed at improving efficiency and organizational performance (D'Adamao et al., 2023; Sushil, 2015). Adequate communication among the stakeholder's is essential and could invaluably foster social sustainability (D'Adamao et al., 2023). Stakeholder engagement and flexibility eventually lead to a resilient and sustainable healthcare system development. Healthcare supply chain (HSC) is far complicated as compared to other supply chains as it includes people and is greatly impacted by the customization of necessary and exact medication supplies

Published online: 13 February 2024

depending on individual patient requirements. It has been observed that to improve flexibility and resilience, application of artificial intelligence (AI) in healthcare sector supply chain is crucial (Vishwakarma et al., 2023).

Salam and Bajaba (2022) elaborated HSCmanagement as the process of handling the movement of medicinal supplies and flow of information among each and every party involved in chain of providing healthcare services. Those included are: suppliers, providers, customers and at last the patients. The industry revolution 4.0 in supply chain is replacing the traditional role of supply chain in health care sector by process digitalization for more flexible and resilient services. COVID-19 pandemic uncovered possible disruptions in healthcare sector (Ishak et al., 2023; Zaoui et al., 2023). For example, the healthcare sector in India suffered a failure in evaluating and making predictions about the medication needs for COVID-19 s wave and in turn, it created major issues among the people and the government of India (Tyagi, 2023). The digitalization of HSC has appeared as the vital phenomena (Priyadarshini et al., 2023) and for its implementation, the organizations are converging themselves on the Omni channel healthcare approach (Bag et al., 2023).

United Nation's 3rd Sustainable Development Goal (SDG) "Affordable healthcare access for all"; looks very challenging considering the population growth (United Nations, 2023). Aging population, shortage of skilled healthcare workers, sudden demand for medicines, lack of capability and resources to manage disruptions are some of the challenges faced by healthcare sector (Gedam et al., 2022; Vishnu et al., 2020). Taking lessons from the dreadful pandemic, its quintessential for healthcare foundations to mandatory reform their support operations by applying emerging technologies such as artificial intelligence, big data, mechanization, analytics, block-chain, and several other pioneering cloud-based solutions for managing the flexibility of organizational resources, businesses, and supply chain resilience (Bali et al., 2023; Kumar et al., 2023; Nguyen et al., 2022; Rahman et al., 2023; Sharma et al., 2023).

Digital technology enhances managerial resilience and process flexibility, which further augments firm's growth (Wided, 2023; Zhang & Qi, 2021). It has been observed that data driven innovations based on emerging technologies such as Artificial Intelligence helps in making process flexible and resilient (Henriksen et al., 2020; Jorzik et al., 2023; Mishra et al., 2022). According to Vishwakarma et al., (2023), Artificial Intelligence (AI) will be instrumental in providing resilient healthcare services to society during disruptions. In general, AI can be defined as the simulation of human intelligence processes by machines, especially digital computers or computer-controlled robots. AI is a powerful tool and technique that makes computers learn and make predictions involving technologies such as machine learning, speech recognition (Abdel-Basset et al., 2021). Niti Ayog (2018) in its report states that "AI is a constellation of technologies that enable machines to act with higher levels of intelligence and emulate the human capabilities of sense, comprehend and act". Digitalization facilitates healthcare networks to attain superior structural and technological outcomes. Application of such digital tools contributes greater coordination between the channels involved, developing better communication network and, subsequently results in improved performance (Jatoba et al., 2023). Globally AI in healthcare market is expected to grow at a compound annual rate of 41.5%. India has a long way to go in AI technology use and is expected to generate 40 million jobs by 2030 (Vijai & Wisetsri, 2021). In India, AI is growing nearly at a very fast rate of 40% per year. Subsequently, witnessing the increased AI use in healthcare, with increased doctor-patient ratio from 4.8:1000 in 2017 to 6.9:1000 by 2023 (Vijai & Wisetsri, 2021). Now many agencies have started exploring the use of AI in various verticals of healthcare. Subsequently, lot of startups are grooming around the world and India is in the top list of nations achieving millions of funding for the tech-based startups for their AI led innovations (Vijai & Wisetsri, 2021). While building AI systems for healthcare, it is essential that they do not contain biases in their algorithms (Henriksen et al., 2020; Marotta, 2022).

In developed countries, AI technologies are being increasingly used in health care sector to solve complicated problems. These nations are successful in employing AI to get feasible and economic growth. They possess good IT infrastructure, optimum resources and advanced technological factors that contributes in developing a smooth healthcare system (Oloyede et al., 2023). On the other hand, in developing countries, AI adaption and implementation is very limited (Polisetty et al., 2023). Although robust program awareness and immense prospective advantages of AI in HSC is there, but the real acceptance rate of this technology is comparatively low. This can be due to the cultural, financial and technological limitations that emerging economies face while implementing healthcare systems smartly (Sreenivasan & Suresh, 2022). However, for developing and resource scarce countries, it is suggested that they should advance and progress toward the application of AI technology in "primary care, service delivery, integration and analysis of medical data, and responses to diseases outbreaks and other medical emergencies" as expressed in the assembly (United Nations, 2023).

AI in healthcare raises several issues pertaining to ethical and social perspective such as on responsibility for the outcomes and errors of AI systems (Chikhaoui et al., 2022). Privacy and security of healthcare data also pose a huge

challenge (Sunarti et al., 2021). This requires a careful consideration and regulation to ensure successful implementation of AI systems. There could be some cultural and behavioral barriers, e.g., the resistance to change, low digital literacy that could subsequently be improved by more awareness (Albahri et al., 2023).

AI can be seen as a game changer in developing countries like India, where the doctor-patient ratio is low along with skewed healthcare infrastructure i.e. more access to urban population (Vaishya et al., 2020). Primarily, AI can be seen most suitable for the repetitive jobs like X-rays, data entry, CT scan, etc. Also, digital nurse is being designed for patient condition monitoring and follow-up and is seen most advantageous because of its 24/7 availability of quick answering for the medication (Haleem et al., 2019; Vijai & Wisetsri, 2021).

In recent past, the AI enabled technology applications are picking up in India. A Bangalore-based start-up has developed a mechanism for screening early signs of breast cancer (Pradhan et al., 2021). Similarly, to overcome the shortage of ophthalmologists in Tamil Nadu, AI is used for detecting diabetic retinopathy. During COVID-19, Government of India employed AI enabled chatbot for ensuring effective communication (Chettri et al., 2020). Similarly, Watson Assistant was deployed by Indian Council of Medical Research (ICMR) for responding to the queries of frontline workers on COVID-19 from various testing and diagnostic facilities in the country. COVID-19 universal epidemic postures a threat to masses of public health and the uncertainty and financial prudence around the world. Only intellect information from varied data sources could deliver the basis for severe medical and social reactions that boost the application of constrained healthcare resources, create customized patient medication plans, aware policy-makers, and speed up clinical trials (Singh et al., 2021, 2023) In the pharmaceutical sector, AI implementation has significantly reduced the time in preclinical drug identification and design process. Additionally, this has also helped in identifying possible pharmaceutical therapies to potentially help in combating the spread of COVID-19 by repurposing drugs (Vijai & Wisetsri, 2021).

AI systems are perceived to be adding value in clinical decision-making, automating repetitive tasks, minimizing errors and creating an effective and efficient utilization of healthcare systems (Pradhan et al., 2021; Vaishya et al., 2020). Contrary to this, the risks are well known and its adoption is challenging considering societal and organizational barriers with the use of technology i.e. algorithmic governance (Janssen & Kuk, 2016). Whereas other barriers such as lack of clarity and standardized guidelines deter AI use in Indian healthcare system. Most of the companies of AI aiding healthcare are startups, which lacks national and international certification. Consecutively, there is less acceptance and trust toward the startups which actually results in hampering and limiting AI implementation in India (Vijai & Wisetsri, 2021).

The field of AI in healthcare is not new and has seen a variety of research work in last few decades. But still this field lacks to portray a wide and complete picture from constitutional, medical and engineering prospective (Petersson et al., 2022). Toorajipour et al. (2021) stated that utmost research studies on AI in the supply chain lacks particular industrial contexts, and their restraining visions may perhaps not be pertinent for other situations. Furthermore, the AI enactment in HSCs has gained a very little attention in the prevailing literature and offers scope for innovative research directions (Kumar et al., 2023). Developing countries like India are struggling to develop technology driven healthcare supply chain to provide efficient, economical and resilient services to its population due to various obstacles. Limited research has been done to explore these challenges in the context of developing economies. In a country like India, AI could provide immense benefits to the healthcare sector through its successful implementation. For a policy drafting and evolving a comprehensive implementation, it is primarily important to understand the barriers and how they impact and what needs to be done to overcome them. In a nutshell, this study attempts to achieve the following research objectives:

RO1. Identifying and prioritizing the barriers of AI adoption in Indian Healthcare sector for flexible and resilient services.
RO2. Analyzing the barriers, their inter-relativity and categorization into cause and effect groups.
RO3. Analyzing social implications of AI applications in healthcare sector.

The rest of the paper are organized as follows. Sect. "Literature Review" summarizes the literature. Sect. "Methodology" identifies the barriers to AI adoption. The methodology is outlined in Sect. "Results and Discussion". Sect. "Managerial and social Implications" contains the results and a discussion. Sect. "Conclusion, Limitations and Future Scope" lists the managerial implications and then conclusion and scope for further research follows in Sect. 7. This study contributes to identifying major barriers to AI application in healthcare sector for flexible and resilient services. Findings will help healthcare professionals in formulating effective strategies for successful implementation of AI to provide efficient and resilient healthcare services to the society in developing countries like India.

Literature Review

This section consists of two sub-sections. First, study explores barriers of AI adoption in healthcare. Then barriers are categorized as cause and effect from strategy perspective.

AI is the capability of a computer program to perform tasks or reasoning processes that we usually associate with intelligence in a human being (Spatharou et al., 2020) or one can define AI as the technology potential to replicate the proficiencies of humans. Since AI does not base its judgments on pre-programmed if–then logic, which yields the same outcomes from the same sets of input instructions, it differs from existing automation technologies. Due to easy access to huge amounts of data, affordable computers, and modern machine learning, AI research witnessed a boom in the first ten years of the twenty-first century. The workforce in the healthcare industry is anticipated to grow by 50% in the next ten years as a result of improved healthcare and increased life expectancy (Spatharou et al., 2020), putting immense pressure on the healthcare system to meet expectations. By enabling practitioners to spend more time providing direct patient care while minimizing burnout, AI in healthcare can boost practitioners' experiences (He et al., 2019). The error rate falls to 0.5% when AI is added to human diagnosis, as compared to 3.5% for human doctors working alone (Begovic et al., 2019). AI can further increase process quality, decrease healthcare outlays and increase convenience (Leone et al., 2021). It has successfully proved itself in achieving patient's overall satisfaction and meeting the expectations showing a better performance as compared to traditional methods (Tani et al., 2022). In its broadest sense, AI employs an interdisciplinary technique to address the challenge of comprehending, formulating, and duplicating the intellect and cognitive acts by utilizing principles and tools from computation, calculation, logic, technicality, and even biology.

The extraordinary opportunities presented by the current advancements in AI in the field of healthcare come with serious drawbacks for patients, developers, providers, and regulators. There is a lack of explainability about these aspects (Amann et al., 2020). AI innovations have the potential to change almost every aspect of health care, from diagnosis, monitoring to data handling (Farouk et al., 2020). The biggest question facing AI in various healthcare fields is not whether the technology will be effective, but rather how to ensure its adoption in routine clinical practice. These difficulties will eventually be resolved, but it will take considerably longer than it will for the technology to advance.

AI is potentially seen as a significant technical enabler to help address challenges of the healthcare to create resilience and to provide potential paths to support businesses on their journey toward long-term sustainability (Modgil et al., 2022). COVID-19 has forced the stakeholders in the healthcare supply chain to embrace AI and enabling technologies for a potential long-term economic sustainability (Ivanov & Dolgui, 2020).

Artificial Intelligence in Healthcare

The Healthcare Supply Chain (HSC) mandates the production and distribution of medicines at appropriate time, at proper place, and in proper quantity (Bag et al., 2021; Beaulieu & Bentahar, 2021; Gupta et al., 2021). Minor negligence can be proved to be brutal and can cause fatal incidents (Brault & Saxena, 2021; Jahan et al., 2021). Healthcare research states use of complex algorithms and software to emulate human intelligence in analysis and interpretation. Biomedical informatics and healthcare planning and policy cannot be separated from AI in healthcare.

The application of AI in healthcare depends upon the integration of various knowledge-based tools rather than standalone consultants. Classification of AI changes considerably with respect to the field, AI in healthcare can be examined through two branches a virtual branch and a physical branch (He et al., 2019). A virtual branch is an informatics approach assisting healthcare practitioners in decision making, helping patients through IOT (internet of things) devices. A physical branch is a directly related approach which includes assistive robots in elderly care and surgery. AI in healthcare is still in early stages and major investment is flowing in from developed countries with the USA expected to account for majority, followed by Western Europe and China (Bohr & Memarzadeh, 2020). In developed countries, AI in healthcare play's a more direct role with implementation in clinical roles whereas in developing countries AI in healthcare is more of a service role providing interface between institute and patients. AI in the initial phase is to address routine, repetitive and large administrative tasks (Spatharou et al., 2020), which reduces time-consuming practitioner's task, elevating operations and expanding reach. In the second phase, AI is expected to support and shift hospital-based care to home based care through virtual assistance, remote monitoring and AI-powered alerts. In the final phase, AI is expected to perform clinical tasks such as surgical treatment and assist individuals in critical decision making.

AI is enhancing the value in the healthcare industry in various ways such as increasing access to data or information, digitization, and a booming startup environment. But the failure of the overall supply chain during Covid-19 disaster that had generated severe paucity of medical supplies and equipments for personal protection, created

distress among caregivers of the healthcare industry. Corona Virus was a catastrophe that knockout the world in the succeeding half of 2019, and has affected the whole world (Arji et al., 2023). According to World Health Organization, since October 1, 2022, this infection had infected over 623,300,000 people taking 6,550,410 lives.

In particular, the COVID-19 outbreak has forced companies to take an active role in driving change both within their organizations and in the upstream and downstream processes of the supply chain. As a result, companies are looking for diversified suppliers who can offer innovative solutions to help them adapt to the changing landscape. To enhance the performance of HSC, prominent researchers have put forward the importance of digitalization techniques that can be successfully applied to HSC (Bali et al., 2023; Kumar et al., 2023). AI-enabled solutions have gained attention in this context, and their relevance in a post-COVID world will only continue to grow. In healthcare, AI tools can help analyze disease patterns, and with the help of machine learning technologies, complex decisions can be made by computers that would be difficult for humans to make alone. This provides a competitive advantage for companies, enabling them to streamline processes, cut costs, and improve efficiency (Vaishya et al., 2020). During the pandemic, AI solutions have been used to deliver primary healthcare via telemedicine, demonstrating their potential to revolutionize healthcare delivery.

Barriers while Implementing AI in Healthcare

An optimum supply chain is cost effective. It reduces risk, enhances quality, swiftness, and reliability (Tarei et al., 2022). It also ensures flexibility and enhances sustainable practices. Still achieving an active supply chain is a challenge for present industries and commerce (Yousefi & Tosarkani, 2022). Most of studies have analysed the benefits of AI technology in supply chain, while some research has also put forward the barriers in adopting this technology (Desingh, 2022). The factors such as safety of patients and health related results create greater complications in HSC, which gets worse by existing multiple partakers that are falling under diverse jurisdictions (Gardas, 2022). Operational matters such as lead time and distractions like calamities and strikes (Ivanov & Dolgui, 2021) pose risks to the supply chain. Besides all these issues, matters such as service disturbances, smaller product life cycle (World Economic Forum, 2019), and variability in climatic conditions also acts as challenging factors for efficient supply chain. Although IoT has contributed at a steady rate, the World Health Organization (WHO) has estimated that even well-developed economies are facing major challenges in healthcare supply chain looking on the ways the present services are provided. Identification of barriers specific to Indian healthcare is more cumbersome than it seems due to India's unique characteristics and position in the world. Swain et al. (2023) explored 15 factors that challenges the efficiency of health application in India. They pointed out that intellectual technologies, safety and maintaining privacy of data are the key hurdles in mHealth adoption.

Identifying and understanding the barriers for AI implementation in Indian healthcare requires understanding of barriers faced by different nations in AI implementation in healthcare and then sub-setting the barriers based on similarities with India specific conditions to truly identify the barriers in AI in healthcare. India possesses vast resources but requires effective governance and also despite having vast pool of IT professionals, it only holds a small share of research in AI in healthcare (Vishnu et al., 2020).

Vuong et al. (2019) revealed the key elements for developing countries in AI implementation in healthcare through a case study on Vietnam Healthcare. Their study highlighted 3 elements – technological, financial sustainability and socio-political commitment as the driving force in nation's readiness for AI implementation.

Alhashmi et al. (2019) investigated use of AI in UAE healthcare system focusing on identifying the key success factors in order to adopt AI in healthcare effectively and efficiently. It is observed that barriers in AI implementation vary with time and it makes the understanding and identification of the barriers more complex. This study identifies barriers faced during AI implementation in Healthcare, specific to Indian conditions. Challenges for developing countries may be different as compared to developed countries (Moktadir et al., 2019).

Through an extensive study of literature review and research in the implementation of AI in healthcare, we have identified 18 barriers that affects directly or indirectly. We discuss these barriers as given below:

C1. *Lack of acceptance and trust of AI among patients:* Countries like the USA, China, Japan, EU countries, etc., have moved further ahead in AI technology to its advanced adaptation stages environment (Alzaharani et al., 2022). However, developing countries like Brazil and India are still in initial adaptation stages. There is a low acceptance of AI technology prevalent among the general public and doctors due to lack of trust toward this new change in the healthcare field (Albahri et al., 2023; Farouk et al., 2020; Petersson et al., 2022; Solanki et al., 2023) It has also been seen that with the increasing introduction of AI there has been degradation of empathic interaction between patients and medical specialists as well as mismatch between existent and expected or needed skills in the workforce (Vishnu et al., 2020).

C2. *Lack of awareness about AI application in healthcare:* Currently in India, AI technology in healthcare is in nascent stage of development and not known by many about its usage. There is a lack of awareness and knowledge among the health practitioners about how to adapt to these new changes in technologies, particularly AI technology in healthcare fields (Arsene, 2019; Gedam et al., 2022; Pradhan et al., 2021).

C3. *Maintaining data privacy of patients:* Data privacy is all about proper handling of data for usage, collection, retention, deletion, and storage. One of the most significant and urgent issues is maintaining data privacy of patients. According to a study, around 93% healthcare establishments of U.S experienced a major security breach and as a result of that they faced several issues such as deny-service attacks, malware, cyber-intrusions, etc. (Ahsan and Siddique, 2022). HSC also faces major challenges namely cyber-security risks such as authentication of data, privacy, data sequencing, and authorization, by adopting IoT. The application of blockchain and data hiding ways can maintain the security of data (Azzaoui et al., 2022). If not dealt properly, it may lead to filing of lawsuits cases against hospitals or AI developers for infringement upon the right to privacy of the general public (Niti Aayog, 2018). In the recent time with the increasing digitization of health records, there has been an increase in the number of cases of data breaches such as leaking medical history of patients, sharing data with AI developers without consent from input providers, and threats from hackers (Farouk et al., 2020). This has resulted in the release of millions of individual's health records (Amann et al., 2020; Pradhan et al., 2021). It has also been concluded in a study that, there will be an increase in such data breaches if the demand is greater than supply of healthcare related data (Albahri et al., 2023; Jiang & Bai, 2020; Neumann et al., 2023; Solanki et al., 2023). Due to the increasing importance of privacy, there are well-defined personal information protection laws, regulations and guidelines in place in countries/regions like the EU which has the General Data Protection Regulation, 2019 (Shabani & Borry, 2018).

C4. *Inadequate regulations for AI application in healthcare:* As AI application in healthcare in India is in nascent stages of adaptation, there hasn't been a need felt previously for having a regulation related to AI in the healthcare sector (NITI Aayog, 2018). Legal and regulatory standards act as a major barrier that influences the acceptance of AI in HSC (Desingh, 2022). However, with the continuous increase in clinical use of AI, there have been emerging potential challenges which need to be monitored by a definite set of regulations (Gedam et al., 2022; Solanki et al., 2023). A legal framework will ensure in maintaining a standard of service. There is also a lack of formal regulation around anonymization of data sets which is being fed to AI systems (Amann et al., 2020; Neumann 2023; Petersson et al., 2022; Pradhan et al., 2021).

C5. *Complexity in managing diversified stakeholders:* There is an involvement of diverse stakeholders in the health sector which includes Government agencies, medical organizations, medical institutions, insurance companies, pharma companies, public and IT firms. These are linked by a complex ecosystem. It results in the absence of collaborative efforts (NITI Aayog, 2018) and a contradiction in mapping challenges between various stakeholders due to the presence of diverse opinions among them.

C6. *Lack of availability of high-quality data:* AI systems require data as the primary source for learning. A large volume of training data is needed by AI algorithms to achieve sufficient performance levels (Albahri et al., 2023; Alzaharani et al., 2022; Farouk et al., 2020; Hsu et al., 2021; Solanki et al., 2023). However, in India, insufficient availability of relevant and robust clinical datasets, has led to failures in the initial adaptation stages (NITI Aayog, 2018). In India, AI developers face problems while developing AI-based products due to the insufficiency in the data pool (Haleem et al., 2019).

C7. *Biased and small data set:* Appropriate selection of training data for AI-based systems is needed because it influences output recommendations, interpretations and observations. Data provided to AI systems from which they get trained always possess a risk of being biased and small set (Albahri et al., 2023). This is due to the environment and people from which these data are being collected and getting AI systems trained, which might not be appropriate for others. For example, data gathered and provided from academic medical colleges for getting AI systems trained, would lead AI systems to know less and treat less effectively than the data gathered and provided from elsewhere. Systematically biased training data like underrepresentation of individuals may lead those biases to be propagated, and scaled in resulting algorithms (Farouk et al., 2020).

C8. *Fear among medical specialists of getting replaced by machines:* With the introduction of electronic health records there are increasing conversion of physicians into data entry clerks, worsening of physician burnout and reducing patient satisfaction. There is increasing fear among medical specialists that with evolving technology and its implementation, their work would be getting replaced by robots and machines especially in the fields of radiology (Albahri et al., 2023; Chockley & Emanuel, 2016; Jha & Topol, 2016; Petersson et al., 2022; Solanki et al., 2023). Also, there has been rising concerns among

few scholars about increasing usage of AI, which would lead to decreasing human knowledge and capacity over time and space, as a result of which practitioners would lose the ability to catch and correct AI errors, which could have further led to augmentation of medical knowledge with AI technology (Cabitza et al., 2017; Froomkin et al., 2019).

In India, the government should build confidence in the initial adaptation stages to each stakeholder including doctors and the general public that AI is to augment rather than replace human intelligence (Vijai & Wisetsri, 2021).

C9. *High adaptation cost of AI and reduced profits for hospitals:* A huge investment is needed in the initial stage for the implementation of AI in healthcare (Hsu et al., 2021; Supeekit et al., 2016). As there is very fewer initial profits, so it lacks interest from market investment (Niti Ayog, 2018). Also, there is an inadequate focus by the Government in its budgetary allocation toward AI implementation in the health sector. The high adaptation cost of AI is due to professionals skilled in areas such as data analytics and AI come at a very high cost than other techies in India (Haleem et al., 2019; Pradhan et al., 2021).

C10. *Lack of understanding on complex AI algorithms:* There have been some of the perceived AI technological related challenges such as complex algorithms designed by IT firms, problems of interoperability, lack of ability to process unstructured data and algorithm opacity to its users in AI systems. The decisions made by the AI system for such a patient treatment plan based on the inputs provided, lacks transparency of AI algorithm to the decision being made (Petersson et al., 2022). This lack of transparency is also known sometimes as "black box" because of not being understandable or modifying nature.

C11. *Fear of injury and risks to patients:* Although a large amount of death occurs due to diagnostic error, which could have been reduced by employing AI in healthcare (Supeekit et al., 2016;). But still there is fear among scientist and medical personnel that if AI systems are employed and if there occur some errors or problems then this could lead to injury or other healthcare-related problems to patients (Albahri et al., 2023; Arsene, 2019; Hsu et al., 2021; Solanki et al., 2023). So, they continue to have an approach of incredible caution.

C12. *Lack of data integration and continuity:* The foundation of AI is based on availability of datasets which should be continued and integrated. Insufficient quantities of availability of data is not a problem since last decade. It has been changed now and it has resulted in transformation of healthcare into heterogeneous data rich environment. Due to heterogeneous nature of data, the datasets lack proper combination, filtration and identification which should have been curated and adequate to feed the system Integrated datasets refer to linkages between personal demographic data (e.g. age, gender) and longitudinal clinical data, including the entire clinical history of patients. Whereas continuity in data can be provided by bringing in fragmented, diverse, uncoordinated data from various or variety of sources such as insurance claims, genetic information, radiological images, electronic health records, care documentation, and laboratory and imaging system records. Further accumulating from various sources into a single place and thereafter processing the data for learning, reasoning, self-correction processes. A major hurdle in data integration and continuity is the lack of definitive laws and regulations for the secondary use of routinely collected data from health records of the general public (Albahri et al., 2023; Neumann et al., 2023).

C13. *Organizational resistance to data sharing:* There is a general lack of will and resistance from hospitals related to sharing of health-related data. This is due to potential misuse of patient data by AI developers and commercial organizations for their personal benefits. Sometimes the misuse of data results in filing of lawsuits against hospitals for sharing of patients' data which lead to infringement upon the right to privacy of the patients (Cohen & Mello, 2019; Neumann et al., 2023). In India, Electronic Health Records (EHR) Policy is put in place, but still sharing of data between various stakeholders and hospitals is in practice. This is because different hospitals which are major providers of patient's data and information had interpreted 'digitizing records' differently (NITI Aayog, 2018).

C14. *Lack of compatibility of machine versus human value judgment:* The decision made by the AI machine is based on data feed and algorithms designed. It lacks human touch and judgments while providing or making decisions. Further there is blindness of emotional cues of patients by machine. In a report it was mentioned that the due to wrong interface of human and AI lead to fatal decision (Vijai & Wisetsri, 2021).

C15. *Lack of accountability:* Decisions made by AI systems are based on a complex network of ecosystems, which is influenced by self-learning AI techniques. As a result of this, there is always a challenge of identifying the accountability for a particular decision being made by a machine (Niti Ayog, 2018). As there are no consequences for being unaccountable there is less incentive for responsible actions (Amann et al., 2020; Petersson et al., 2022).

C16. *Lack of research on AI applications in healthcare:* There is an inadequate well-developed startup ecosystem for transforming the core research into market

application. In India, there is limited core research related to AI technologies due to insufficient incentives for doing research in this field (Haleem et al., 2019; Niti Ayog, 2018).

C17. *Inadequate availability of AI expertise, skilled manpower and skilling opportunity:* In comparison to other developed nations, India lacks sufficient expertise and manpower in AI technologies. Attributing to this fact, only 4% of AI professionals in India have experience with revolutionary technologies like deep learning and neural networks. This indicates the need of capacity building with a gradual investment in skills and AI expertise (NITI Aayog, 2018; Pradhan et al., 2021; Vishnu et al., 2020; Vijai & Wisetsri, 2021).

C18. *Lack of explainability:* The operation of AI based technology is based on black-box system, as it does not provide reasoning and transparency behind a particular decision, classification or forecast being made by the system. So as a result, it lacks acceptance toward its decision from doctors and the general public (Amann et al., 2020; Vijai & Wisetsri, 2021; Vuong et al., 2019). These barriers are summarized in Table 1.

Several studies around the globe state use of AI in different capabilities in healthcare sector and the challenges are evident for its adoption in India too (Vishwakarma et al., 2023). To assess the identified barriers utilizing the DEMATEL approach, we employ a thorough research methodology mentioned in the next section.

Methodology

Following a thorough assessment of the literature and research, we identified 18 main barriers in adoption of AI in health care in India. In the current context, we prioritize them and solve them in order to ensure the smooth application of AI in the healthcare industry. Figure 1 depicts the methods used to identify, prioritize, and categorize barriers to the deployment of AI in the health care sector. The research is divided into two stages. The first phase involves identifying barriers through a literature research and discussions with subject experts. In the second phase, barriers are prioritized and classified using DEMATEL technique.

We considered five experts, who have sufficient knowledge and experience in the field of AI and healthcare, as well as the ability to provide consistent and rational judgments. The basis for selection of experts was based on some criteria, such as: academic qualification specifically in AI, healthcare or related disciplines; a professional experience of at least 5 years in a related discipline. Details of the experts is given in Table 2.

DEMATEL is a structural modeling tool that is commonly used to analyze the causal link of complex structures using matrices or diagraphs. It is a particularly powerful tool for identifying cause-and-effect relationships in a complicated system. DEMATEL is used to confirm dependency among cause-and-effect group components using matrices (James et al., 2023). And it may also be used to create an impact relation map, which shows the relative link between the components and can be utilized to investigate and solve complex problems.

Although there are several MCDM techniques that are used for a particular problem set. In such techniques, AHP analyzes the distribution of a goal among the elements and compares and judges which element has greater impact on that goal. The VIKOR method formulates a ranking index based on a specific measure of "closeness" to the ideal solution by using linear normalization, and the Best Worst Method provides pairwise comparison. However, DEMATEL is favored over the other approaches in this scenario because it efficiently analyzes the mutual influence of numerous aspects and understands complicated cause-and-effect relationships in decision-making problems (Hsu et al., 2021; James et al., 2023).

Several researchers have applied the DEMATEL methodology in various healthcare-related studies. Supeekit et al. (2016) identified several supply chain performance groups (e.g. clinical care process efficiency) and cost, time, reliability, and productivity as its aspects using DEMATEL and ANP methodologies together. Ghadami et al. (2021) used the DEMATEL approach for evaluating standards of hospital accreditation and stated that management and leadership play an influential role in their improvement. Vishnu et al. (2020) identified various operational risk factors in Indian healthcare and used the DEMATEL approach along with others to investigate their causal relationships. Alzahrani et al. (2022) proposed several service quality factors that support the use of mHealth, a service platform during COVID-19.

Causal relationships among these factors were identified using DEMATEL as these factors help support the continued use of the platform during COVID-19.

Gedam et al. (2022) used DEMATEL for identifying critical success factors in the Indian healthcare industry. Hsu et al. (2021) explored various trends (IoT, Blockchain, Telemedicine) developing in the healthcare industry and analyzed their mutually influential relationship utilizing a combination of variants of DEMATEL methodology. These studies highlight the versatility and effectiveness of DEMATEL in analyzing complex relationships and contributing to the improvement of healthcare services.

A detailed stepwise procedure of DEMATEL methodology is given as follows:

Step 1. Gathering expert opinion.

Table 1 Barriers to application of AI in Health Sector

Code	Barrier name	References
C1	Lack of acceptance and trust by patient	Albahri et al. (2023); Alzaharani et al. (2022); Farouk et al. (2020); Petersson et al. (2022); Solanki et al. (2023); Vishnu et al. (2020)
C2	Lack of awareness about AI technology application in healthcare	Arsene (2019); Gedam et al. (2022); Jorzik et al. (2023); Pradhan et al. (2021); Vishwakarma et al. (2023)
C3	Maintaining Data Privacy of patients	Albahri et al. (2023); Amann et al. (2020); Farouk et al. (2020); Jiang and Bai (2020); Neumann et al. (2023); NitiAyog (2018); Pradhan et al. (2021); Shabani and Borry (2018); Solanki et al. (2023)
C4	Inadequate Regulations for AI application in healthcare	Amann et al. (2020); Desingh (2022); Gedam et al. (2022); Neumann et al. (2023), NitiAyog (2018); Petersson et al. (2022); Pradhan et al. (2021); Solanki et al. (2023)
C5	Complexity in managing diversified stakeholders	Niti Ayog (2018); Vishwakarma et al. (2023)
C6	Lack of Availability of high-quality data in standardized formats	Albahri et al. (2023); Alzaharani et al. (2022); Farouk et al. (2020); Haleem et al. (2019); Hsu et al. (2021); Niti Ayog (2018); Solanki et al. (2023)
C7	Possibility of data being Biased and Unequal	Farouk et al. (2020); Albahri et al. (2023), Vishwakarma et al. (2023)
C8	Fear among medical specialists of getting replaced by machine	Cabitza et al. (2017); Chockley and Emanuel (2016); Froomkin et al. (2019); Jha and Topol (2016); Petersson et al. (2022); Vijai and Wisetsri (2021)
C9	High adaptation cost of AI and reduced profits for hospitals	Haleem et al. (2019); Hsu et al. (2021); Niti Ayog (2018); Pradhan et al. (2021); Supeekit et al. (2016)
C10	Lack of understanding on complex AI algorithms	Petersson et al. (2022); Vishwakarma et al. (2023)
C11	Fear of injury and risks to Patients	Arsene (2019); Hsu et al. (2021); Solanki et al. (2023); Supeekit et al. (2016)
C12	Lack of data integration and continuity	Albahri et al. (2023); Neumann et al. (2023)
C13	Organizational resistance to data sharing	Cohen and Mello (2019); Neumann et al. (2023); Niti Ayog (2018)
C14	Lack of compatibility of machine versus human value judgement	Vijai and Wisetsri (2021)
C15	Lack of Accountability	Amann et al. (2020); Niti Ayog (2018); Petersson et al. (2022)
C16	Lack of research on AI applications in healthcare	Haleem et al. (2019); Niti Ayog (2018)
C17	Inadequate availability of AI expertise, manpower and skilling opportunity	Niti Ayog (2018); Pradhan et al. (2021); Vijai and Wisetsri (2021); Vishnu et al. (2020)
C18	Lack of Explainability	Amann et al. (2020); Vijai and Wisetsri (2021); Vuong et al. (2019)

Initially, the literature review is performed and expert opinion are sought to identify the factors. Further, data is collected through expert's responses on a predesigned questionnaire to form an initial direct relation matrix while rating each factor on the basis of its influence on the other factor. The direct influence between any two factors is rated by experts on the bases of scale given as, 0–'No influence'; 1–'Little influence'; 2–'High influence'; 3–'Very high influence'.

Step 2. Constructing initial direct relation matrix (A).

The matrix 'A' is created based on the expert responses. The individual response shows the extent to which the responder believes factor p influences factor q. The diagonal elements are set to zero for $p = q$. If K is the number of respondents and 'n' is the number of factors with $1 \leq K \leq H$, then for each respondent a $(n \times n)$ non-negative matrices is established as $X^k = [x^k_{pq}]$. It is possible to createX^1,X^2, X^3 …, X^H matrices based on H respondents, respectively (H denotes the number of experts). Now, 'A' can be created by averaging the responses of all respondents and given by Eq. (1.

$$A = [a_{pq}] \tag{1}$$

Where, 'a' is the element of matrix, A as expressed in Eq. (2)

$$a_{pq} = \frac{1}{H}\sum_{K=1}^{H} * x^k_{pq} \tag{2}$$

Step 3. Calculating the normalized initial direct-relation matrix (D).

Matrix 'D' is created by normalization performed according to Eq. 3.

$$D = S \times A \tag{3}$$

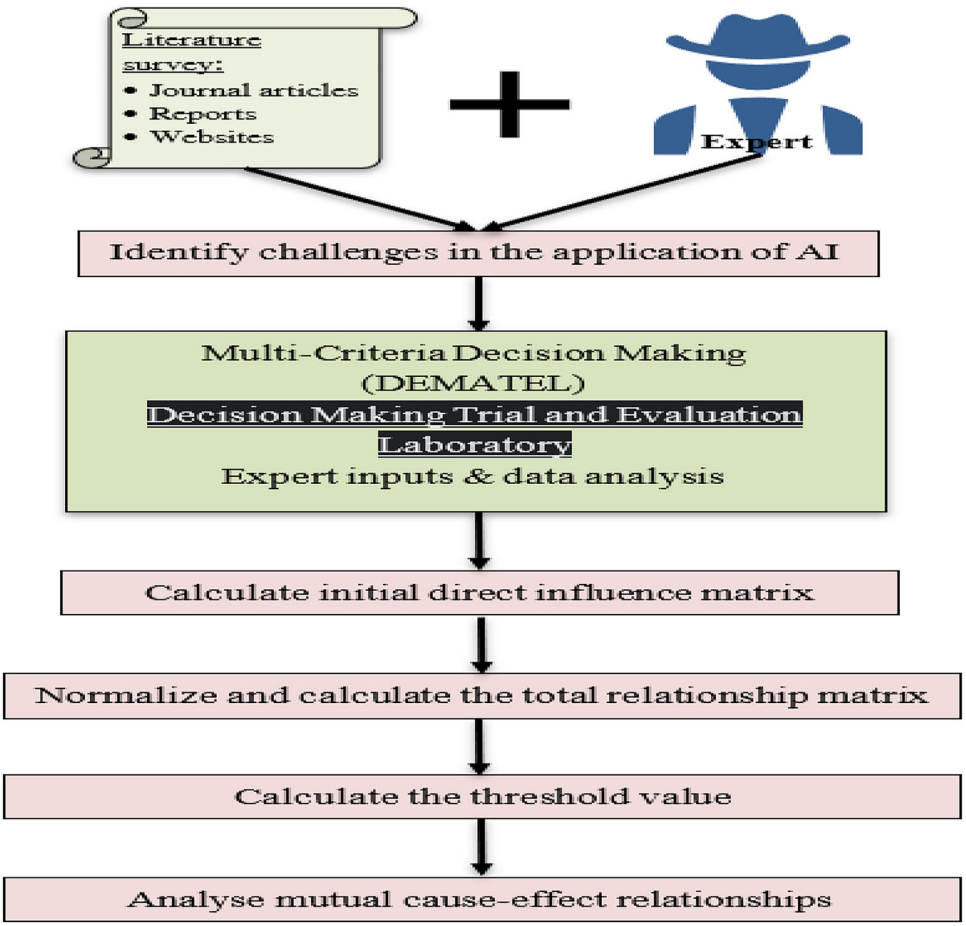


Fig. 1 Research methodology

Table 2 Profile of the experts

Sr. No	Expert designation	Expertise and field exposure
1	Chief Medical Officer (Clinician)	Expertise in healthcare administration and policy-making with 14 years of experience
2	AI in healthcare researcher (Industry)	Doctorate degree in Artificial Intelligence applications in healthcare and data analytics with 6 years of experience
3	Health Informatics specialist (Industry)	Masters degree with experience in implementing health information systems and AI tools with 10 years of experience
4	Public health expert (Regulatory body)	Masters in public health research and policy; expertise in healthcare systems with an experience of approx. 10 years
5	Data Privacy and Security (Industry)	Specialists dealing with healthcare data, cybersecurity and privacy issues

$$\text{Where} S = \frac{1}{max\sum_{q=a}^{n} * a_{pq}}, 1 \le \text{p} \le \text{n} \quad (4)$$

Step 4. Deriving the total relation matrix (T).

The matrix 'T' is constructed by the following equation

$$T = D(I - D)^{-1} \quad (5)$$

Here, 'I' denotes Identity matrix.

$$T = \left[t_{pq}\right]_{n\times n} p,\ q = 1,\ 2, \ldots, n \quad (6)$$

Step 5. Calculating the sum of rows (R) and columns (C) of matrix T.

Let R be vector indicating sum of rows of matrix 'T''.

$$R = \left[\sum_{q=1}^{n} * t_{pq}\right]_{n\times 1} = [tp]n \times 1 \quad (7)$$

and.

let C be 1 vector indicating sum of columns of matrix 'T'

$$C = \left[\sum_{p=1}^{n} * t_{pq}\right]_{1\times n} = [tq]1 \times n \quad (8)$$

Step 6. Calculate the threshold value in order to create a cause and effect relationship diagram.

In a cause and effect diagram, a data set is represented by (Ri + Ci, Ri − Ci), where (Ri + Ci) is represented on a horizontal axis (also known as 'Prominence'), which ranks the factors. Whereas (Ri − Ci) is represented on a vertical axis (also known as 'Relation'), which indicates the category to which the factor belongs. A positive value of (Ri − Ci) states the cause category while the negative value states the effect category of the factors. The threshold value is calculated by averaging all factors in the matrix 'T' and then constructing a casual digraph. In matrix 'T', the overall value represents how a factor influences the other. Only values greater than the threshold value, are used in constructing relationship maps.

Results and Discussion

The barriers in application of AI in Healthcare sector are selected through extensive literature review and responses obtained from experts through a list of questions targeted on the identified barriers. Total 18 barriers are listed in Table 1. Each barrier is weighted relative to other barriers by experts based on their influence on a scale of 0–4 as per Table 3. The responses from experts are compiled in a pair-wise matrix to assess the impact of each barrier. Based on Eq. (1) and (2), an average of values from responses is combined into one to yield an average direct relation matrix, given Table 4. The next process is to normalize the initial average direct relation matrix using the Eqs. (3) and (4). This yields in the formation of a Normalized Direct Relation Matrix which is shown in Table 5. The total relation matrix is obtained from the normalized direct relation matrix by using Eq. (5). The total relation matrix is presented as Table 6. The arithmetic mean of all factors in the total relation matrix (T) is used as the threshold value. The threshold value expresses how a particular factor influences the others (Alzahrani et al., 2022). In this study-

Threshold Value = 0.1603.

To understand the structural links among variables, the threshold value (0.1603) was set to exclude the insignificant factors in matrix T. By summing the means of the components in the total matrix, we were able to get the total threshold value.

Using Eq. (7) and (8), value of Ri (sum of row) and Ci (sum of column) of total relationship matrix is calculated. The value of (Ri + Ci) and (Ri − Ci) yields cause-effect table which indicates whether a factor is a cause or is an effect as mentioned in Table 7.

A causal digraph (as given in Fig. 2) is formed by putting (Ri + Ci) values in the Abscissa Axis (x-axis) and (Ri − Ci) values in the Ordinate Axis (y-axis). The digraph depicts the factor's interaction relationship and splits them into cause-and-effect groups. The causal effect map aids in understanding the structure by distinguishing the influence of one factor over another and filtering out insignificant effects. The (Ri + Ci) value aids in assessing the relative relevance of various factors. Higher the magnitude of (Ri + Ci) value i.e. 7.390 for Inadequate Regulations for AI application in healthcare, higher will be the importance of that factor. The factors that have a negative (Ri − Ci) value are considered in the effect group i.e. − 0.889 for High adaptation cost of AI and reduced profits for hospitals and are the ones that are being influenced by other factors. Whereas, factors that have a positive (Ri − Ci) value are put in the cause group (0.583 for Lack of awareness about AI technology application in healthcare) and are the ones that influence other factors. The (Ri − Ci) value determines how much a factor influences other factors. The greater the size of the positive (Ri − Ci) value, the greater its influence on other factors.

An inter-relatability is formed among 18 barriers for application of AI in healthcare. The factors are clubbed in cause and effect section based on the data analyzed through DEMATEL. The cause group are the dominant one, which has impact on other factors and include factors such as Inadequate regulations for AI application in healthcare, lack of awareness about AI technology application in healthcare, Inadequate availability of AI expertise, man-power and skilling opportunity, lack of compatibility of machine versus human value Judgment, lack of data integration and continuity, lack of explainability, lack of availability of high quality data in standardized formats, Fear among medical specialists of getting replaced by machine, lack of understanding on complex AI algorithms, maintaining data privacy of patients.

Table 3 Influence weightage

Influence category	Weightage
No Influence	0
Low Influence	1
Medium Influence	2
High Influence	3
Very High Influence	4

Table 4 Initial direct relation matrix

Cr	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
C1	0	2	1	2	1	0.7	2	1.7	2.3	1	2.7	0.7	1.7	2	2	0.7	2	3
C2	3.3	0	1.3	2.3	2.3	1.7	1.3	1.7	2.7	2.3	2.7	1.3	1.3	2.7	1.7	1.7	2	1.7
C3	2.7	1.3	0	2.7	2	1.3	1.7	1.3	0.3	0.7	1.3	2	2.7	0.7	1.3	1.7	0.7	0.3
C4	2.3	1.7	3.7	0	1.7	3	1.3	2.7	1.7	1	3.3	2	3.3	2	2.7	2.3	1.7	1.3
C5	0.7	1	1	3	0	1.3	2.3	2	2.3	1.7	0.7	2.7	1.7	1	2.7	1.3	2	1.3
C6	0	2	2	1	1.3	0	2.7	0.7	1.7	1	0.3	4	1.7	2	1.3	1.7	1.3	1.7
C7	1.7	0	0.7	2.3	1	2.3	0	0	2.7	0.7	2.3	1.3	0.3	3.3	2.3	0.7	1	1.3
C8	1	0.7	1.3	3	2.3	0.7	1.3	0	2.7	1.7	1.7	0.7	2	1.7	1.3	1.7	1	1
C9	0.3	2.3	1.7	2	2.3	1	0.3	2	0	1	1.3	1	2.7	0.7	1.3	1.7	3.3	0
C10	2	1	0.3	1.3	2.7	1	1.3	1.3	1	0	1.3	1	1	2.3	2	2.3	2	3.7
C11	1.7	1	0	1.7	1	1	1	2.3	1.7	1	0	0.3	0.7	2.7	2	2.3	1.7	1.3
C12	1.3	1.7	1.7	2	2	3	1.7	0.7	1.7	0.3	1.7	0	2.3	1.7	1.7	2.7	2.3	1.3
C13	1.3	0.3	3	2	1.7	1.7	1.3	0	2.3	0.7	0.3	3	0	0.3	1	1.3	1.3	1.3
C14	3	2.7	0	2	2	1.7	1.3	3	2	1.3	1.7	0.7	0.7	0	1	2	2	2.3
C15	2.7	2	0.3	3	2.3	0.7	1.7	1.7	0.3	1	2.7	0	0.3	1.7	0	1	1.7	2.3
C16	1.7	2.7	0.7	2	1	1	2.3	1.3	3.7	1.3	1.7	2.3	1.3	1	1.7	0	2	1
C17	1	3.3	1.3	1.3	2.7	1.7	1	0.7	3	2	2.7	1.7	1.3	2.3	2	2.3	0	1.7
C18	3	2	2	1.7	1.7	1.7	1	1	2	1.3	2.7	0.7	0.7	1.3	0.7	2.7	1.7	0

Table 5 Normalized direct matrix

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
C1	0	0.05	0.03	0.05	0.03	0.02	0.05	0.04	0.06	0.03	0.07	0.02	0.04	0.05	0.05	0.02	0.05	0.08
C2	0.09	0	0.04	0.06	0.06	0.04	0.04	0.04	0.07	0.06	0.07	0.04	0.04	0.07	0.04	0.04	0.05	0.04
C3	0.07	0.04	0	0.07	0.05	0.04	0.04	0.04	0.01	0.02	0.04	0.05	0.07	0.02	0.04	0.04	0.02	0.01
C4	0.06	0.04	0.1	0	0.04	0.08	0.04	0.07	0.04	0.03	0.09	0.05	0.09	0.05	0.07	0.06	0.04	0.04
C5	0.02	0.03	0.03	0.08	0	0.04	0.06	0.05	0.06	0.04	0.02	0.07	0.04	0.03	0.07	0.04	0.05	0.04
C6	0	0.05	0.05	0.03	0.04	0	0.07	0.02	0.04	0.03	0.01	0.11	0.04	0.05	0.04	0.04	0.04	0.04
C7	0.04	0	0.02	0.06	0.03	0.06	0	0	0.07	0.02	0.06	0.04	0.01	0.09	0.06	0.02	0.03	0.04
C8	0.03	0.02	0.04	0.08	0.06	0.02	0.04	0	0.07	0.04	0.04	0.02	0.05	0.04	0.04	0.04	0.03	0.03
C9	0.01	0.06	0.04	0.05	0.06	0.03	0.01	0.05	0	0.03	0.04	0.03	0.07	0.02	0.04	0.04	0.09	0
C10	0.05	0.03	0.01	0.04	0.07	0.03	0.04	0.04	0.03	0	0.04	0.03	0.03	0.06	0.05	0.06	0.05	0.1
C11	0.04	0.03	0	0.04	0.03	0.03	0.03	0.06	0.04	0.03	0	0.01	0.02	0.07	0.05	0.06	0.04	0.04
C12	0.04	0.04	0.04	0.05	0.05	0.08	0.04	0.02	0.04	0.01	0.04	0	0.06	0.04	0.04	0.07	0.06	0.04
C13	0.04	0.01	0.08	0.05	0.04	0.04	0.04	0	0.06	0.02	0.01	0.08	0	0.01	0.03	0.04	0.04	0.04
C14	0.08	0.07	0	0.05	0.05	0.04	0.04	0.08	0.05	0.04	0.04	0.02	0.02	0	0.03	0.05	0.05	0.06
C15	0.07	0.05	0.01	0.08	0.06	0.02	0.04	0.04	0.01	0.03	0.07	0	0.01	0.04	0	0.03	0.04	0.06
C16	0.04	0.07	0.02	0.05	0.03	0.03	0.06	0.04	0.1	0.04	0.04	0.06	0.04	0.03	0.04	0	0.05	0.03
C17	0.03	0.09	0.04	0.04	0.07	0.04	0.03	0.02	0.08	0.05	0.07	0.04	0.04	0.06	0.05	0.06	0	0.04
C18	0.08	0.05	0.05	0.04	0.04	0.04	0.03	0.03	0.05	0.04	0.07	0.02	0.02	0.04	0.02	0.07	0.04	0

The effect group are impacted by the factors on cause group and include factors such as High adaptation cost of AI and reduced profits for hospitals, complexity in managing diversified stakeholders, lack of research on AI applications in healthcare, lack of acceptance and trust by patients, fear of injury and risks to patients, lack of accountability, possibility of data being biased and unequal, organizational resistance to data sharing.

The above results are in line with the previous researches in this field. The challenge- inadequate regulations for

Table 6 Total relation matrix

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
C1	0.13	0.17	0.12	0.2	0.16	0.13	0.16	0.15	0.21	0.12	0.21	0.12	0.15	0.18	0.18	0.15	0.18	0.19
C2	0.24	0.15	0.15	0.24	0.22	0.17	0.17	0.18	0.25	0.17	0.23	0.16	0.17	0.22	0.19	0.2	0.21	0.18
C3	0.18	0.14	0.09	0.2	0.17	0.13	0.14	0.13	0.14	0.09	0.15	0.15	0.17	0.13	0.15	0.16	0.13	0.11
C4	0.22	0.2	0.22	0.2	0.21	0.22	0.18	0.21	0.23	0.14	0.26	0.2	0.23	0.21	0.23	0.23	0.21	0.18
C5	0.15	0.15	0.13	0.23	0.14	0.15	0.17	0.16	0.21	0.13	0.16	0.18	0.16	0.16	0.2	0.17	0.18	0.15
C6	0.12	0.17	0.14	0.17	0.16	0.11	0.17	0.12	0.18	0.11	0.14	0.21	0.15	0.17	0.15	0.17	0.16	0.15
C7	0.15	0.11	0.1	0.19	0.14	0.15	0.09	0.1	0.19	0.09	0.18	0.13	0.11	0.19	0.17	0.13	0.14	0.13
C8	0.14	0.13	0.13	0.22	0.18	0.12	0.13	0.1	0.2	0.12	0.17	0.12	0.16	0.16	0.15	0.16	0.15	0.13
C9	0.12	0.17	0.13	0.19	0.18	0.13	0.11	0.15	0.14	0.11	0.16	0.13	0.17	0.13	0.15	0.16	0.2	0.1
C10	0.18	0.15	0.1	0.18	0.2	0.13	0.14	0.14	0.18	0.09	0.17	0.13	0.13	0.18	0.17	0.19	0.18	0.21
C11	0.15	0.13	0.08	0.17	0.14	0.12	0.12	0.15	0.17	0.1	0.12	0.1	0.11	0.17	0.16	0.17	0.15	0.13
C12	0.17	0.17	0.15	0.21	0.19	0.19	0.16	0.13	0.2	0.1	0.18	0.12	0.18	0.18	0.17	0.2	0.2	0.15
C13	0.14	0.11	0.16	0.18	0.15	0.14	0.13	0.09	0.18	0.09	0.12	0.17	0.1	0.11	0.13	0.14	0.14	0.12
C14	0.21	0.2	0.1	0.21	0.19	0.16	0.15	0.19	0.21	0.13	0.19	0.13	0.14	0.14	0.16	0.19	0.19	0.18
C15	0.19	0.16	0.1	0.22	0.18	0.12	0.14	0.15	0.15	0.11	0.2	0.1	0.11	0.16	0.12	0.15	0.16	0.17
C16	0.17	0.19	0.12	0.21	0.16	0.14	0.17	0.15	0.24	0.13	0.18	0.17	0.15	0.16	0.17	0.13	0.19	0.14
C17	0.17	0.22	0.14	0.21	0.22	0.17	0.15	0.14	0.24	0.15	0.22	0.17	0.16	0.2	0.19	0.21	0.15	0.17
C18	0.2	0.17	0.15	0.19	0.17	0.15	0.14	0.14	0.2	0.12	0.2	0.13	0.13	0.16	0.14	0.19	0.17	0.11

Table 7 Categorization of factors-Cause & Effect

Sn	Ri	Ci	Ri + Ci	Ri—Ci	Category
C1	3.77997069	3.611	7.39087069	0.169070689	Cause
C2	3.49784709	2.914	6.41224709	0.583447085	Cause
C3	3.30538799	3.103	6.40798799	0.20278799	Cause
C4	3.06738345	3.013	6.08018345	0.054583454	Cause
C5	3.06258385	2.609	5.67118385	0.453983845	Cause
C6	2.86984129	2.715	5.58444129	0.155241294	Cause
C7	2.72946988	2.633	5.36246988	0.096469884	Cause
C8	2.66289572	2.569	5.23159572	0.094195718	Cause
C9	2.87742862	2.105	4.98252862	0.772328624	Cause
C10	2.56971204	2.32	4.88931204	0.250112044	Cause
C11	2.62804922	3.517	6.14524922	-0.88915078	Effect
C12	2.97709321	3.166	6.14319321	-0.18900679	Effect
C13	2.97625887	3.084	6.06035887	-0.10784113	Effect
C14	2.91876903	3.034	5.95286903	-0.11533097	Effect
C15	2.44526671	3.233	5.67846671	-0.78793329	Effect
C16	2.68225451	2.984	5.66665451	-0.3021455	Effect
C17	2.49174391	2.636	5.12764391	-0.14415609	Effect
C18	2.39256222	2.689	5.08196222	-0.29683779	Effect

AI application in healthcare- was present in nearly every paper we reviewed. In our results as well, it ranked highest in the cause-effect matrix and signifies a maximum impact on other factors. Though another study present a bias and contradiction in different stakeholders' view and that's what our results show "the complexity in managing diversified stakeholders" in effect group but still ranks higher which represents its overall impact in successful adoption of AI in healthcare. In another work, Vuong et al. (2019) presents technological, financial and socio-political

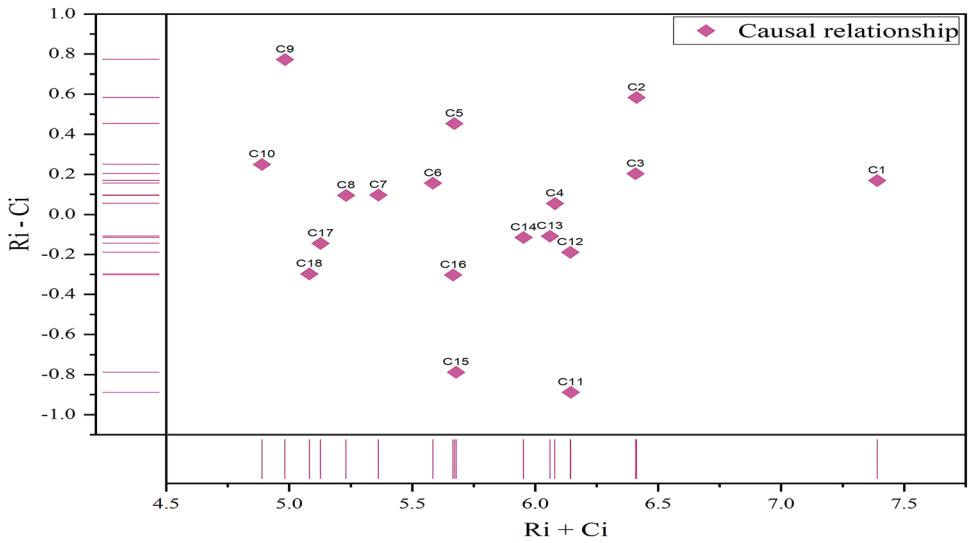


Fig. 2 Causal and effect groups of barriers

as 3 driving forces in the realization of the change; our data also signifies the same by placing these challenges in the order such as Lack of data integration and continuity, High adaptation cost of AI and reduced profits for hospitals and Lack of awareness about AI technology application in healthcare in the cause group.

The research signifies a broad exploration of the barriers along with a potential to engage further with the multiple stakeholders. This eventually states how this paper covers a broader scope of AI applications and stakeholders in health care, and how our barriers could address or mitigate the challenges faced by the digital health innovators. The DEMATEL method adds value to the analysis by prioritizing and evaluating the interdependence among the barriers. Furthermore, indicating a scope of work, how our Indian context differs from the international/regional context in terms of the demand and supply of AI skills, and how barriers and enablers could affect the labour market and the diffusion of AI in health care.

Developing and implementing a comprehensive and coherent regulatory framework for AI in healthcare that addresses the ethical, legal, and social issues such as privacy. The regulatory framework should be based on the principles of fairness, accountability, and transparency, and should involve the participation and consultation of various stakeholders, such as healthcare providers, policy makers.

Understanding of AI in healthcare among the healthcare professionals and awareness in the general public should be increased which will foster trust and acceptance of AI systems. This could be done by providing effective communication and education strategies, such as campaigns, workshops. Ai expertise access and skilling opportunities should be increased. Also, there is a need to adopt common frameworks, guidelines for data collection.

Managerial and Social Implications

This section has two subsections as given below.

Managerial Implications

The research has several managerial implications. First, the study attempts to fill the gap by identifying barriers in adopting AI application for a more resilient healthcare. The study integrates the domain specific knowledge, barriers derived from the literature and also portrayed in the context of COVID-19. The study demonstrates a comprehensive view of the identified barriers studying their mutual relationships. The barriers are characterized in terms of the ability of medical professionals (users) to determine if specific actions can be completed by a given technology from the standpoint of usability. To understand the usability a practitioner's view was considered and the feasibility of the considered factors in the real-life settings. The influential causal relationship between the barriers is assessed to determine the significant challenges in the adoption of the AI in healthcare in India.

The methodology can be used to a wide range of disciplines to solve various academic and practical difficulties, such as the one presented in this study. The findings indicate specific barriers concerning application of AI in Indian healthcare sector and pays way to further the research in the area. Therefore, it is a valuable contribution to the extent body of research. Dominant causal factors include Inadequate Regulations for AI application in healthcare, inadequate availability of skilled manpower, lack of compatibility considering machines vs humans, and data related aspects. Whereas, the significant challenge is also due to high adaptation cost and the hospitals do not see a profitability in it. Subsequently, need to manage the

stakeholders in the entire chain for the successful adaptation.

Based on the results of this study, this research will inspire further studies to evolve implementation frameworks and mixed methods to study the scenario wise possible aftereffects. This paper successfully demonstrates the foundation work and developing an understanding on what barriers are down the line and their causal relationships. In summary, the proposed work can a guiding document or study for implementation of AI in Indian healthcare sector.

Social Implications

The study has several social implications specially for the healthcare sector in developing countries like India. The foremost challenge is to manage the patient load and better allocate the resources to ensure efficient healthcare delivery. Thus, removing the barriers through policy level interventions can pave way toward a resilient and sustainable healthcare system. AI application can help provide healthcare access to remote areas in India, especially first in diagnostics and preventive healthcare. This could cope up the shortage of doctors by providing efficient diagnostics, treatment plans and ensure better management of already resource scarce sector. Awareness campaigns and training are required to build trust and acceptance of the AI systems in the population. Subsequently, India need to build capacity and skilled workforce to use the AI for healthcare, ensure transparency and accountability, given the population diversity and the varied healthcare needs. Subsequently, the data privacy and protection of personal information is a widely discussed topic and need to ensure adequate regulations to safeguard patient information.

In a nutshell, the AI application in Indian healthcare sector has long term rewards but need to address the above discussed barriers for successful implementation.

Conclusion, Limitations and Future Scope

In conclusion, while AI holds promise for enhancing healthcare efficiency, its nascent implementation in India requires meticulous attention from researchers and practitioners. This paper contributes by identifying 18 major barriers and employing DEMATEL for prioritization. Recognizing the importance of flexibility in overcoming barriers, the study guides healthcare stakeholders toward successful AI implementation, emphasizing the urgent need to address regulatory gaps, enhance awareness, and foster AI expertise.

The analysis suggests that inadequate regulations, awareness gaps, and shortage of AI expertise present critical barriers to successful AI implementation in healthcare sector. Main contributions of this study can be summarized as follows:

This study identified some of the major barriers to the application of AI in healthcare field emphasizing need for flexibility in health services to enhance resilience.
This proposed methodology provides a flexible framework for prioritizing and understanding causal relationships among barriers enabling adaptability in addressing evolving challenges.
This study will also be assisting Hospitals, Industry related to AI and Government toward right decisions and enable them to identify barriers that could be faced during AI implementation.

The outcomes of this study will assist policymakers, managers, and researchers working in this field in developing a framework for successful AI application in the healthcare industry for efficient service to society. AI implementation is viewed not only as a trillion-dollar enterprise for the future, but also as a means of boosting the efficiency and resiliency of India's overburdened healthcare system. As a result, the healthcare sector should take proactive steps in this direction to avoid roadblocks caused by barriers to AI application in India.

The scope and generalizability of this study are limited by the number and diversity of the experts and the barriers that are considered. Only 18 barriers were selected based on the literature review and the experts' opinion, which may not cover all the possible barriers that exist or emerge in the field of AI and healthcare in India. The collected data from the experts may be subject to bias, error, or inconsistency due to the subjective and qualitative nature of the data. Additionally, the implications and recommendations of this study are based on the assumptions and scenarios considered.

Further, since the sector specific barriers could be different and thus the findings of the current study cannot be generalized to other sectors. It will be interesting to see how the AI adaptation in other sectors is different e.g. education, agriculture, finance etc. Like any of the study, this too has some limitations. One is that the limited number of experts i.e. fifteen experts in this case; were involved. The second is, approaches other than DEMATEL could be utilized and the results could further be validated through multiple case studies.

Acknowledgements Authors would like to express sincere thanks to the Editor and reviewers for valuable inputs and suggestions to improve quality and content of this paper.

Funding This research received no funding and was performed by the authors independently.

Declarations

Conflict of interest There is no conflict of interest.

Ethical Approval The study does not include animal/human data. The used data is anonymized and does not breach any confidentiality/ethical issue.

References

Abdel-Basset, M., Chang, V., & Nabeeh, N. A. (2021). An intelligent framework using disruptive technologies for COVID-19 analysis. *Technological Forecasting and Social Change, 163*, 120431.

Abu Zwaida, T., Elaroudi, K., & Beauregard, Y. (2021). The challenges of drug shortages in the Canadian hospital pharmacy supply chain—a systematic literature review. *Journal of Public Health*. https://doi.org/10.1007/s10389-021-01485-2

Ahsan, M. M., Luna, S. A., & Siddique, Z. (2022). Machine-learning-based disease diagnosis: A comprehensive review. *Healthcare, 10*(3), 541.

Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., & Deveci, M. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Information Fusion, 96*, 156–191.

Alhashmi, S. F., Salloum, S. A., & Abdallah, S. (2019). Critical success factors for implementing artificial intelligence (AI) projects in Dubai Government United Arab Emirates (UAE) health sector: applying the extended technology acceptance model (TAM). In *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics*, pp. 393–405, Springer International Publishing.

Alzahrani, A. I., Al-Samarraie, H., Eldenfria, A., Dodoo, J. E., & Alalwan, N. (2022). Users' intention to continue using mHealth services: A DEMATEL approach during the COVID-19 pandemic. *Technology in Society, 68*, 101862.

Amann, J., Blasimme, A., Vayena, E., Frey, D., Madai, V. I., Precise4Q Consortium. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making, 20*, 1–9.

Arji, G., Ahmadi, H., Avazpoor, P., & Hemmat, M. (2023). Identifying resilience strategies for disruption management in the healthcare supply chain during COVID-19 by digital innovations: A systematic literature review. *Informatics in Medicine Unlocked, 38*, 101199.

Arsene, C. (2019). Artificial Intelligence in Healthcare: the future is amazing. Healthcare Weekly, Source: https://healthcareweekly.com/artificial-intelligence-in-healthcare/. Accessed on 7 Dec 2022.

Ayog, N. (2018). National Strategy for artificial intelligence, Discussion paper. Source: https://indiaai.gov.in/documents/pdf/NationalStrategy-for-AI-Discussion-Paper.pdf, Accessed on 21 Nov 2021.

Azzaoui, A. E., Sharma, P. K., & Park, J. H. (2022). Blockchain-based delegated Quantum Cloud architecture for medical big data security. *Journal of Network and Computer Applications, 198*, 103304.

Bag, S., Dhamija, P., Singh, R. K., Rahman, M. S., & Sreedharan, V. R. (2023). Big data analytics and artificial intelligence technologies based collaborative platform empowering absorptive capacity in health care supply chain: An empirical study. *Journal of Business Research., 154*, 113315.

Bag, S., Gupta, S., Choi, T. M., & Kumar, A. (2021). Roles of innovation leadership on using big data analytics to establish resilient healthcare supply chains to combat the COVID-19 pandemic: A multimethodological study. *IEEE Transactions on Engineering Management*. https://doi.org/10.1109/TEM.2021.3101590

Bali, A. S., & Ramesh, M. (2023). Knowledge–practice gap in healthcare payments: The role of policy capacity. *Policy and Society, 42*(3), 406–418.

Beaulieu, M., & Bentahar, O. (2021). Digitalization of the healthcare supply chain: A roadmap to generate benefits and effectively support healthcare delivery. *Technological Forecasting and Social Change, 167*, 120717.

Begovic, M., Oprunenco, A., & Sadiku, L. (2019). *Let's Talk about Artificial Intelligence*. UNDP.

Bohr, A., & Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. *Artificial Intelligence in Healthcare*. https://doi.org/10.1016/B978-0-12-818438-7.00002-2

Brault, N., & Saxena, M. (2021). For a critical appraisal of artificial intelligence in healthcare: The problem of bias in mHealth. *Journal of Evaluation in Clinical Practice, 27*(3), 513–519.

Cabitza, F., Rasoini, R., & Gensini, G. F. (2017). Unintended consequences of machine learning in medicine. *JAMA, 318*(6), 517–518.

Chettri, S., Debnath, D., & Devi, P. (2020). Leveraging digital tools and technologies to alleviate COVID-19 pandemic. *Available at SSRN 3626092*.

Chikhaoui, E., Alajmi, A., & Larabi-Marie-Sainte, S. (2022). Artificial intelligence applications in healthcare sector: Ethical and legal challenges. *Emerging Science Journal, 6*(4), 717–738.

Chockley, K., & Emanuel, E. (2016). The end of radiology? Three threats to the future practice of radiology. *Journal of the American College of Radiology, 13*(12), 1415–1420.

Cohen, I. G., & Mello, M. M. (2019). Big data, big tech, and protecting patient privacy. *JAMA, 322*(12), 1141–1142.

D'Adamo, I., Gastaldi, M., Piccioni, J., & Rosa, P. (2023). The role of automotive flexibility in supporting the diffusion of sustainable mobility initiatives: A stakeholder attitudes assessment. *Global Journal of Flexible Systems Management, 24*(3), 459–481.

De la Gala-Velásquez, B., Hurtado-Palomino, A., & Arredondo-Salas, A. Y. (2023). Organisational flexibility and innovation performance: The moderating role of management support. *Global Journal of Flexible Systems Management*, *24*(2), 219–234.

Desingh, V. (2022). Internet of Things adoption barriers in the Indian healthcare supply chain: An ISM-fuzzy MICMAC approach. *The International Journal of Health Planning and Management, 37*(1), 318–351.

Dhar, B. K., Stasi, A., Döpping, J. O., Gazi, M. A. I., Shaturaev, J., & Sarkar, S. M. (2022). Mediating role of strategic flexibility between leadership styles on strategic execution: A study on Bangladeshi private enterprises. *Global Journal of Flexible Systems Management*, *23*(3), 409–420.

Farouk, A., Alahmadi, A., Ghose, S., & Mashatan, A. (2020). Blockchain platform for industrial healthcare: Vision and future opportunities. *Computer Communications, 154*, 223–235.

Froomkin, A. M., Kerr, I., & Pineau, J. (2019). When AIs outperform doctors: Confronting the challenges of a tort-induced over-reliance on machine learning. *Ariz. l. Rev., 61*, 33.

Gardas, B. B. (2022). Organizational hindrances to Healthcare 4.0 adoption: An multicriteria decision analysis framework. *Journal of Multi-Criteria Decision Analysis, 29*(1–2), 186–195.

Gedam, V., Raut, R., Inamdar, Z., Narkhede, B., Dharaskar, S., & Narvane, V. (2022). COVID-19 critical success factors in Indian healthcare industry—A DEMATEL approach. *Journal of Multi-Criteria Decision Analysis, 29*(1–2), 135–149.

Ghadami, L., Masoudi Asl, I., Hessam, S., & Modiri, M. (2021). Developing hospital accreditation standards: Applying fuzzy

DEMATEL. *International Journal of Healthcare Management, 14*(3), 847–855.
Gupta, H., Kumar, S., Kusi-Sarpong, S., Jabbour, C. J. C., & Agyemang, M. (2021). Enablers to supply chain performance on the basis of digitization technologies. *Industrial Management & Data Systems, 121*(9), 1915–1938.
Haleem, A., Javaid, M., & Khan, I. H. (2019). Current status and applications of artificial intelligence (AI) in medical field: An overview. *Current Medicine Research and Practice, 9*(6), 231–237.
He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine, 25*(1), 30–36.
Henriksen, A., & Bechmann, A. (2020). Building truths in AI: Making predictive algorithms doable in healthcare. *Information, Communication & Society, 23*(6), 802–816.
Hsu, W. C. J., Liou, J. J., & Lo, H. W. (2021). A group decision-making approach for exploring trends in the development of the healthcare industry in Taiwan. *Decision Support Systems, 141*, 113447.
Ishak, S., Shaharudin, M. R., Salim, N. A. M., Zainoddin, A. I., & Deng, Z. (2023). The effect of supply chain adaptive strategies during the COVID-19 pandemic on firm performance in Malaysia's semiconductor industries. *Global Journal of Flexible Systems Management*, *24*(3), 439–458.
Ivanov, D., & Dolgui, A. (2020). Viability of intertwined supply networks: extending the supply chain resilience angles towards survivability. A position paper motivated by COVID-19 outbreak. *International Journal of Production Research, 58*(10), 2904–2915.
Ivanov, D., & Dolgui, A. (2021). A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0. *Production Planning & Control, 32*(9), 775–788.
Jahan, I., Ullah, I., Griffiths, M. D., & Mamun, M. A. (2021). COVID-19 suicide and its causative factors among the healthcare professionals: Case study evidence from press reports. *Perspectives in Psychiatric Care, 57*(4), 1707–1711.
James, A. T., Asjad, M., Kumar, G., Shukla, V. C., & Arya, V. (2023). Analyzing barriers for implementing new vehicle scrap policy in India. *Transportation Research Part d: Transport and Environment, 114*, 103568.
Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly, 33*(3), 371–377.
Jatobá, M. N., Ferreira, J. J., Fernandes, P. O., & Teixeira, J. P. (2023). Intelligent human resources for the adoption of artificial intelligence: A systematic literature review. *Journal of Organizational Change Management., 36*(7), 1099–1124.
Jha, S., & Topol, E. J. (2016). Adapting to artificial intelligence: Radiologists and pathologists as information specialists. *JAMA, 316*(22), 2353–2354.
Jiang, J., & Bai, G. (2020). Types of information compromised in breaches of protected health information. *Annals of Internal Medicine, 172*(2), 159–160.
Jorzik, P., Yigit, A., Kanbach, D. K., Kraus, S., & Dabić, M. (2023). Artificial intelligence-enabled business model innovation: Competencies and roles of top management. *IEEE Transactions on Engineering Management*. https://doi.org/10.1109/TEM.2023.3275643
Kumar, A., Mani, V., Jain, V., Gupta, H., & Venkatesh, V. G. (2023). Managing healthcare supply chain through artificial intelligence (AI): A study of critical success factors. *Computers & Industrial Engineering., 175*, 108815.
Leone, D., Schiavone, F., Appio, F. P., & Chiao, B. (2021). How does artificial intelligence enable and enhance value co-creation in industrial markets? An exploratory case study in the healthcare ecosystem. *Journal of Business Research, 129*, 849–859.
Marotta, A. (2022). When ai is wrong: Addressing liability challenges in women's healthcare. *Journal of Computer Information Systems, 62*(6), 1310–1319.
Mishra, R., Singh, R. K., & Papadopoulos, T. (2022). Linking digital orientation and data-driven innovations: A SAP-LAP linkages framework and research propositions. *IEEE Transactions on Engineering Management., 71*, 1346–1358.
Modgil, S., Singh, R. K., & Hannibal, C. (2022). Artificial intelligence for supply chain resilience: Learning from Covid-19. *The International Journal of Logistics Management, 33*(4), 1246–1268.
Moktadir, M. A., Ali, S. M., Paul, S. K., & Shukla, N. (2019). Barriers to big data analytics in manufacturing supply chains: A case study from Bangladesh. *Computers & Industrial Engineering, 128*, 1063–1075.
Neumann, O., Guirguis, K., & Steiner, R. (2023). Exploring artificial intelligence adoption in public organizations: a comparative case study. *Public Management Review*. https://doi.org/10.1080/14719037.2022.2048685
Nguyen, D. C., Pham, Q. V., Pathirana, P. N., Ding, M., Seneviratne, A., Lin, Z., & Hwang, W. J. (2022). Federated learning for smart healthcare: A survey. *ACM Computing Surveys (CSUR), 55*(3), 1–37.
Oloyede, A., Fark, N., Noma, N., & Tebep, E. (2023). Measuring the impact of the digital economy in developing countries: A systematic review and meta-analysis. *Available at SSRN 4106167*.
Petersson, L., Larsson, I., Nygren, J. M., Nilsen, P., Neher, M., Reed, J. E., & Svedberg, P. (2022). Challenges to implementing artificial intelligence in healthcare: A qualitative interview study with healthcare leaders in Sweden. *BMC Health Services Research, 22*(1), 1–16.
Polisetty, A., Chakraborty, D., Kar, A. K., & Pahari, S. (2023). What determines AI adoption in companies? Mixed-method evidence. *Journal of Computer Information Systems*. https://doi.org/10.1080/08874417.2023.2219668
Pradhan, K., John, P., & Sandhu, N. (2021). Use of artificial intelligence in healthcare delivery in India. *J Hosp Manag Health Policy, 5*, 1–10.
Priyadarshini, J., Singh, R. K., Mishra, R., & Dora, M. (2023). Application of additive manufacturing for a sustainable healthcare sector: Mapping current research and establishing future research agenda. *Technological Forecasting and Social Change, 194*, 122686.
Rahman, A., Hossain, M. S., Muhammad, G., Kundu, D., Debnath, T., Rahman, M., & Band, S. S. (2023). Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges and open issues. *Cluster Computing, 26*(4), 2271–2311.
Salam, M. A., & Bajaba, S. (2022). Corporate social responsibility during the COVID-19 pandemic: A sequential mediation analysis. *Social Responsibility Journal, 18*(6), 1188–1208.
Shabani, M., & Borry, P. (2018). Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. *European Journal of Human Genetics, 26*(2), 149–156.
Sharma, B., Mittal, M. L., Soni, G., & Ramtiyal, B. (2023). An implementation framework for resiliency assessment in a supply chain. *Global Journal of Flexible Systems Management*, *24*(4), 591–614.
Singh, A., Sushil, P., & Sharma, H. K. (2023). Total interpretive structural modeling-polarity (TISM-P) to analyze the impact of energy on the sustainability performance of hotels: a case study. *Environment, Development and Sustainability,*. https://doi.org/10.1007/s10668-023-03485-6

Singh, S., Dhir, S., Evans, S., & Sushil. (2021). The trajectory of two decades of global journal of flexible systems management and flexibility research: A bibliometric analysis. *Global Journal of Flexible Systems Management, 22*, 377–401.

Solanki, P., Grundy, J., & Hussain, W. (2023). Operationalising ethics in artificial intelligence for healthcare: A framework for AI developers. *AI and Ethics, 3*(1), 223–240.

Spatharou, A., Hieronimus, S., & Jenkins, J. (2020). Transforming healthcare with AI: The impact on the workforce and organisations, source: https://www.mckinsey.com/industries/healthcare/our-insights/transforming-healthcare-with-ai. Accessed on 5 Dec 2020.

Sreenivasan, A., & Suresh, M. (2022). Future of healthcare start-ups in the era of digitalization: Bibliometric analysis. *International Journal of Industrial Engineering and Operations Management, 4*(1/2), 1–18.

Sunarti, S., Rahman, F. F., Naufal, M., Risky, M., Febriyanto, K., & Masnina, R. (2021). Artificial intelligence in healthcare: Opportunities and risk for future. *Gaceta Sanitaria, 35*, S67–S70.

Supeekit, T., Somboonwiwat, T., & Kritchanchai, D. (2016). DEMATEL-modified ANP to evaluate internal hospital supply chain performance. *Computers & Industrial Engineering, 102*, 318–330.

Sushil. (2015). Strategic flexibility: The evolving paradigm of strategic management. *Global Journal of Flexible Systems Management, 16*(2), 113–114.

Swain, S., Muduli, K., Kumar, A., & Luthra, S. (2023). Analysis of barriers of mHealth adoption in the context of sustainable operational practices in health care supply chains. *International Journal of Industrial Engineering and Operations Management*. https://doi.org/10.1108/IJIEOM-12-2022-0067

Tani, M., Troise, C., De Bernardi, P., & Han, T. (2022). Innovating the supply chain in health-related crises: Some evidence from ISINNOVA case. *European Journal of Innovation Management, 25*(6), 716–734.

Tarei, P. K., Kumar, G., & Ramkumar, M. (2022). A Mean-Variance robust model to minimize operational risk and supply chain cost under aleatory uncertainty: A real-life case application in petroleum supply chain. *Computers & Industrial Engineering, 166*, 107949.

Toorajipour, R., Sohrabpour, V., Nazarpour, A., Oghazi, P., & Fischl, M. (2021). Artificial intelligence in supply chain management: A systematic literature review. *Journal of Business Research, 122*, 502–517.

Tyagi, S. (2023). Analytics in healthcare supply chain management in the new normal era: a review and future research agenda. *Benchmarking: an International Journal.* https://doi.org/10.1108/BIJ-03-2023-0155

United Nations (2023). Sustainable development goals. Source: https://sdgs.un.org/goals/goal3; Accessed on 7 Jan, 2023.

Vaishya, R., Javaid, M., Khan, I. H., & Haleem, A. (2020). Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 14*(4), 337–339.

Vijai, C., & Wisetsri, W. (2021). Rise of artificial intelligence in healthcare startups in India. *Advances in Management, 14*(1), 48–52.

Vishnu, C. R., Sridharan, R., Ram Kumar, P. N., & Regi Kumar, V. (2020). Analysis of the operational risk factors in public hospitals in an Indian state: A hybrid DEMATEL–ISM–PROMETHEE approach. *International Journal of Health Care Quality Assurance, 33*(1), 67–88.

Vishwakarma, L. P., Singh, R. K., Mishra, R., & Kumari, A. (2023). Application of artificial intelligence for resilient and sustainable healthcare system: systematic literature review and future research directions. *International Journal of Production Research*. https://doi.org/10.1080/00207543.2023.2188101

Vuong, Q. H., Ho, M. T., Vuong, T. T., La, V. P., Ho, M. T., Nghiem, K. C. P., & Ho, R. C. (2019). Artificial intelligence vs. natural stupidity: Evaluating AI readiness for the vietnamese medical information system. *Journal of Clinical Medicine, 8*(2), 168.

Wided, R. (2023). IT capabilities, strategic flexibility and organizational resilience in SMEs Post-COVID-19: A mediating and moderating role of big data analytics capabilities. *Global Journal of Flexible Systems Management, 24*(1), 123–142.

World Economic Forum, W. (2019). *The global risks report*. World Economic Forum.

Yousefi, S., & Tosarkani, B. M. (2022). An analytical approach for evaluating the impact of blockchain technology on sustainable supply chain performance. *International Journal of Production Economics, 246*, 108429.

Zaoui, S., Foguem, C., Tchuente, D., Fosso-Wamba, S., & Kamsu-Foguem, B. (2023). The viability of supply chains with interpretable learning systems: The case of COVID-19 vaccine deliveries. *Global Journal of Flexible Systems Management, 24*(4), 633–657.

Zhang, J., & Qi, L. (2021). Crisis preparedness of healthcare manufacturing firms during the COVID-19 outbreak: Digitalization and servitization. *International Journal of Environmental Research and Public Health, 18*(10), 5456.

Key Questions Addressed through this Research are on

I. What are major barriers for AI adoption in Indian healthcare sector for flexible and resilient services
II. How these barriers are interrelated. Can we categorise them from strategy perspective
III. How AI applications in healthcare sector may impact service quality of healthcare operations for society

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.



Girish Kumar is a Professor in Mechanical Engineering Department at Delhi Technological University, Delhi. He has more than 25 years of experience in industry, teaching, and research. He holds a PhD degree in Reliability Engineering from Indian Institute of Technology Delhi, India. His teaching and research areas include Optimization, Reliability Engineering, Maintenance Management, Quality engineering, Stochastic Modeling, etc. Dr. Kumar has published more than 70 papers in International Journals and conference proceedings.



Rajesh Kr Singh is Professor in Operations Management Area and Chairperson (Admissions) at Management Development Institute (MDI), Gurgaon, India. His total google citations is around 10000 and h index 51. Currently he is serving as Department Editor to IEEE Transactions on Engg Management (ABDC-AS-ABS-3*) Associate Editor to International Journal of Consumer Studies (ABDC- "A" category journal, Wiley publication), International Journal of Global Business and Competitiveness (Springer publication) and Area Editor, Operations Management Research. He has also figured among the top 2% researchers in the world as per study done by Stanford University scientists continuously from 2020-2023. He has published papers in highly reputed and impactful journals such as Technological Forecasting and Social Change, Business Strategy and the Environment, Resources, Conservation and Recycling, Journal of Cleaner Production, Production planning and Control, Annals of Operations Research, International Journal of Production Research, Journal of Business Research, Applied Soft Computing, etc.



Vedpal Arya is a research fellow in the Department of Industrial Systems Engineering and Management at the National University of Singapore (NUS). His research interests are in digital twin, maritime, logistics, resilience, performance evaluation, supply chain management with diverse applications. After his masters and PhD from IIT Delhi, he worked with the Centre for Next Generation Logistics at NUS, primarily working in industrial collaborative settings focused on warehousing and logistics. He has been a visiting researcher at the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland.



Shivam Kumar Mishra is an undergraduate student in Mech Engg Dept at Delhi Technological University, Delhi, India. His area of interest is artificial intelligence and machine learning applications.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/378359143

View Online Export Citation CrossMark Analyzing the performance of class-F power amplifier by comparing microstrip and SSL/SIL transmission lines based matching networks for sm...

Article · January 2024

DOI: 10.1063/5.0192543

CITATIONS
0

READS
17

2 authors, including:



Puja Priya
Delhi Technological University
6 PUBLICATIONS **2** CITATIONS

SEE PROFILE

All content following this page was uploaded by Puja Priya on 21 February 2024.

The user has requested enhancement of the downloaded file.

AIP Conference Proceedings

RESEARCH ARTICLE | FEBRUARY 20 2024

Analyzing the performance of class-F power amplifier by comparing microstrip and SSL/SIL transmission lines based matching networks for smart parking solutions for improved efficiency and output power FREE

Puja Priya ✉; Gurjit Kaur

Check for updates

AIP Conf. Proc. 2986, 030064 (2024)
https://doi.org/10.1063/5.0192543

View Online Export Citation CrossMark

21 February 2024 13:04:41



AIP Publishing

Analyzing the Performance of Class-F Power Amplifier by Comparing Microstrip and SSL/SIL Transmission Lines Based Matching Networks for Smart Parking Solutions for Improved Efficiency and Output Power

Puja Priya[a], and Gurjit Kaur[b]

Department of Electronics and Communication Engineering, Delhi Technological University, Bawana Rd, Delhi-110042, India

[a]Corresponding author: puja_2k19phdec502@dtu.ac.in
[b]gurjitkaur@dtu.ac.in

Abstract. A cascade Class-F power amplifier (PA) is proposed for smart parking solutions. Input and output matching network are designed and analyzed using Microstrip line (MSL), suspended shielded lines (SSL) and shielded inverted line (SIL). Class-F PA design is based on the harmonic termination of an even and odd harmonic for improved efficiency. Class-F PA produces non-overlapping drain current of half sinusoidal form and voltage waveforms of half square form to reduce power dissipation. Suspended shielded (SSL) and shielded inverted lines (SIL) have low attenuation, low propagation loss, low insertion loss (S11) and low dielectric constant as compared to Microstrip line (MSL). The main objective of the work is to compare the transmission line like MSL, SSL and SIL for better performance of Class-F PA in terms of high efficiency and bandwidth. Design is based on even and odd harmonic termination to get non-overlapping current and voltage waveform. Simulation is performed using advanced design system (ADS) software using gallium nitride high electron mobility transistor (GaN HEMT) device from CREE. The results show improved power added efficiency (PAE) of 80 to 85%, gain of 15dB and output power (Pout) of 44dBm with suspended shielded (SSL) and shielded inverted (SIL) line as compared to Microstrip line (MSL) at 0 to 3GHz frequency.

Keywords*: Class-F, Transmission lines, ADS, GaN HEMT, SSL/SIL***.**

21 February 2024 13:04:41

INTRODUCTION

Due to sudden population growth and urban development traffice congestion in the cities is increased. Smart parkingis the only solution for this problem [1].Internet of things (IoT) will resolve this problem by connecting multiple physical devices and forms the inter- network of the devices [2]. Smart parking architeture consists of smart sensors that remain connected to the internet, each parking slot determine the occupied and empty parking space by sensing the presence of the car (object). Sensors collect the signal and process it, then transmit the information using smart gateway to the cloud[3][4]. This cloud will store the information/data for the various other customers that use smart parking application in the portable device like mobile phones as shown in Fig 1. The transmitter of each sensor consist of power amplifier (PA), whose main aim is to boost the strength of the signal and drive the antenna. Class-F and inverse Class-F power amplifiers are highly efficient power amplifiers. These amplifiers uses harmonic termination topologies to enhance efficiency and output power. In Class-F, the device output voltage are designed into square wave and current waveform is designed into half-sinusoidal waveform. These waveforms are obtained by matching at operational frequencies and harmonics, where as the inverse Class-F has output voltage waveform as half-sinusoidal and current waveform as half-squared[5]–[7].

3rd International Conference on Functional Materials, Manufacturing, and Performances
AIP Conf. Proc. 2986, 030064-1–030064-6; https://doi.org/10.1063/5.0192543
Published by AIP Publishing. 978-0-7354-4840-7/$30.00

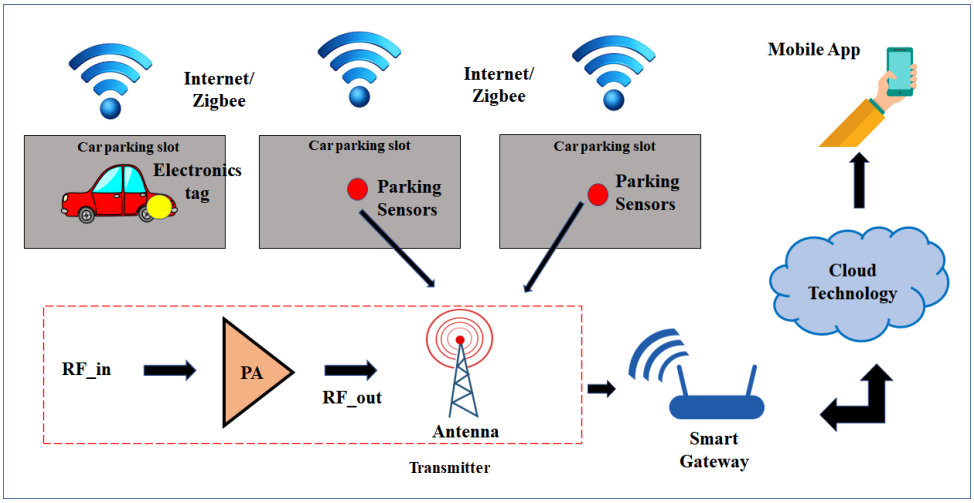


FIGURE 1. Showing Smart Parking architecture based on IoT and Cloud Technique using Power Amplifier in the Transmitter

From previous work many Class-F and Inverse Class-F has been designed and simulated. Mostly design obtained familiar outcomes like the gain, the power added efficiency (PAE) and the output power (Pout). In [6] a Class-F PA was designed at 0.8 to 2.7GHz frequency using GaN HEMT technology, where compensation circuit was used for matching impedances. Step-impedance (harmonic matching network) HMN method was used to terminate 3rd harmonic and bandwidth enlargement. Simulated results show 20-50 W of saturated output power, 52 to 76% of drain efficiency, and gain more than 10dB. In [7], inverse Class-F PA was designed for healthcare applications using 10W GaN technology where cylindrical cavity resonator was designed. Measured results represent drain efficiency of 70%, output power of 40dBm, and 10dB gain over operating frequency. In [8], designed a multi-octave Class-F PA at 0.2 to 2.5GHz using GaN HEMT technology, where simulated results represents drain efficiency of 55.5% to 70.3%, output power of 43.7 to 46.9dBm. In [9], designed a Class-F PA at 2.4GHz with novel harmonic control circuit which contains input matching network (IMN), output matching network (OMN) and lampshade compact microstrip resonant cell (LCMRC). The measured results show PAE of 83% and output power of 23.65dBm. In [10], designed a Class-F PA for multi-octave bandwidth. Measured results represent wide bandwidth of 128.5%, drain efficiency of 60%, output power of 39.2dBm and gain of 11.7dB. In [12] designed an inverse Class-F PA in stacked topology, using 130nm SiGe BiCMOS technology at 38GHz. This design contains harmonic matching network and double stacked power stages. The measured results show output power of 21.2dBm, gain of 22.1dB, and PAE of 30.1%. In [13] designed a Class-F PA for extended bandwidth, using 10W GaN technology at 2.6GHz. Short circuited small characteristics impedance transmission line was used in the harmonic control network to obtain extended bandwidth. Measured results show drain efficiency of 74.37%, output power of 40.82 dBm, and gain of 11.82dBm. In[14], designed a Class-F PA for non-linearity and broadband analysis. The proposed PA was designed using GaN technology at 1.75 to 2.3GHz. The measured results show output power of 53.2dBm, drain efficiency more than 65%, and ACLR of -52dBc with instantaneous bandwidth of 395MHz. In [15], an inverse Class-F PA using GaN pHEMT technology at 2.4GHz. a novel HCN was designed to lower the effects of fundamental frequency reflection, harmonics termination, enhancement of PAE and to obtain non-overlapping voltage and current waveforms.

This paper analyzed a Class-F PA design using Microstrip line (MSL), shielded lines (SSL) and shielded inverted line (SIL) and compared the results. The paper is classified into following section: part II will discuss circuit implemented, part III represents results and part IV will provide conclusion of the work.

CIRCUIT IMPLEMENTED

Cascade Class-F Power Amplifiers

Proposed cascade Class-F PA design is shown in the Fig 2, which contains input-matching network, output matching network, stability network and bias tees. CGH40010F GaN HEMT device from CREE is utilized to design PA with Vds=28V and Ids = 200mA.Cascade topology is used to obtain high gain and efficiency in PA. This topology produces till 5th harmonics and their load impedances can be obtain from load pull simulation. Microstrip

21 February 2024 13:04:41

transmission line (MSL), suspended shielded transmission lines (SSL) and shielded inverted transmission line (SIL) are used to analyze and compare the performance of PA. MTL are the most common and easily fabricated transmission line (TL), SSL and SIL are another type of TL which shows various advantages over MSL. SSL and SIL have low attenuation, small dielectric constant, low propagation loss and low insertion loss [16]. Inductor is replaced by the MSL, SSL and SIL to see the behavior of the circuit. SSL and SIL are able to obtain large bandwidth. R1 and C3 are the stability network, this will improve the stability factor (K>1) and make the device unconditionally stable over the desired frequency band. Capacitors C1 and C2 are the blocking capacitor that shows high impedance for the DC signal and low impedance for the RF signal. Capacitors C4 and C5 are the bypass capacitors that bypasses the high frequency component to the ground. Input Matching Network (IMN) and Output Matching Network (OMN) are designed and compared using MSL, SSL and SIL. Smith chart utility tool is used for designing of matching network. IMN is designed to avoid gain variation of the device and to get maximum power gain. Output matching network is designed which can terminate 2nd harmonic. 2ndharmonic is the most crucial harmonic in the PA design for determining its good efficiency. The design is analyzed using continuous wave (CW) signal.

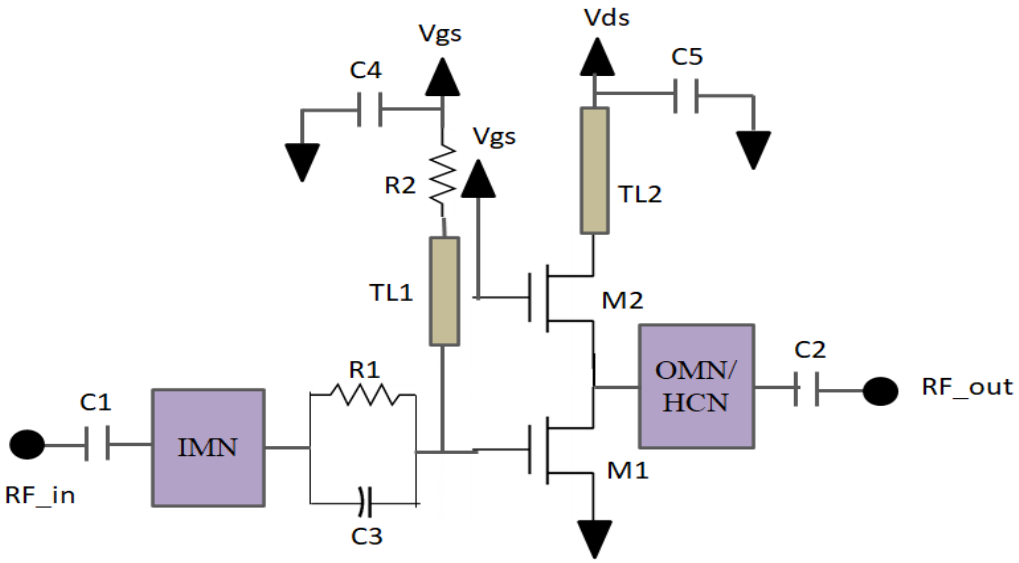


FIGURE 2. Cascade Class-F Power Amplifier

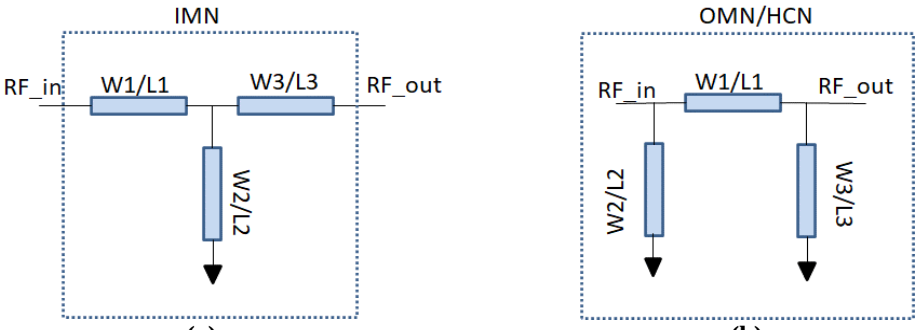


FIGURE 3. (a) Input matching network (IMN), **(b)** output matching network (OMN)

RESULTS

Even and odd harmonic analysis is performed to identify the performance of Class-F PA. Design and simulation of the circuit is done by "Advance Design System (ADS) Software by Keysight. S-parameter analysis is performed using S-simulator by terminating input and output port of the design at 50ohm. Fig 4 (a) showing return loss (S11)

21 February 2024 13:04:41

analysis from 2 to 3 GHz. From the result it is visible that highest return loss(S11) is obtained at 2.3GHz which is more than -35dB. SSL and SIL represents better results in comparison to MSL.

In fig 5. Output power (Pout) vs Input power (Pin) analysis is shown by comparing the MSL, SSL and SIL performance. From 0 to 13dBm the circuit shows linear behavior. At 20dBm circuit operate at saturation. Again, SSL and SIL show better Pout (more than 40dBm) in comparison to MSL. In fig 6 Pout versus power added efficiency (PAE) is analyzed, obtained PAE is more than 80% at 44dBm output power. Again, SSL and SIL shows better results in comparison to MSL. Fig 7 shows the forward gain (S21) which between 14 to 16dB at 0 to 3GHz. SSL and SIL shows highest forward gain.

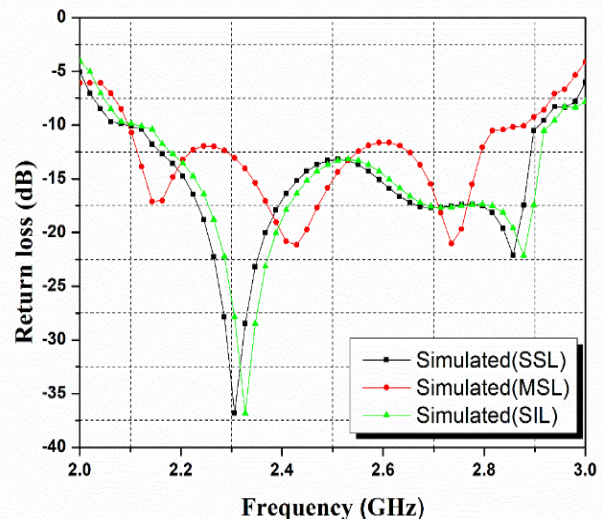


FIGURE 4. Return loss comparison using MSL, SSL and SIL

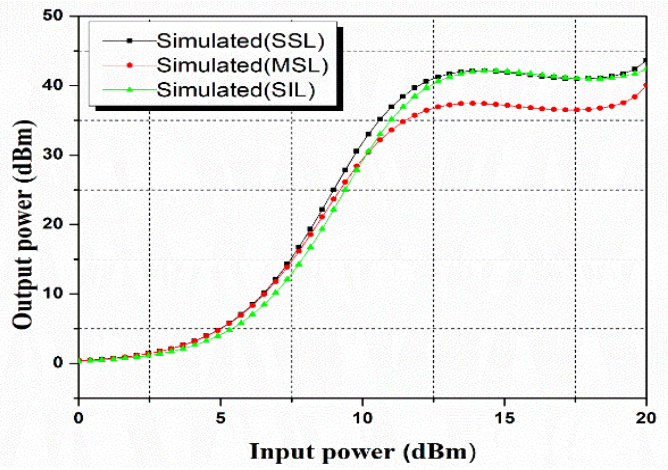


FIGURE 5. Output Power Vs Input Power comparison using MSL, SSL and SIL

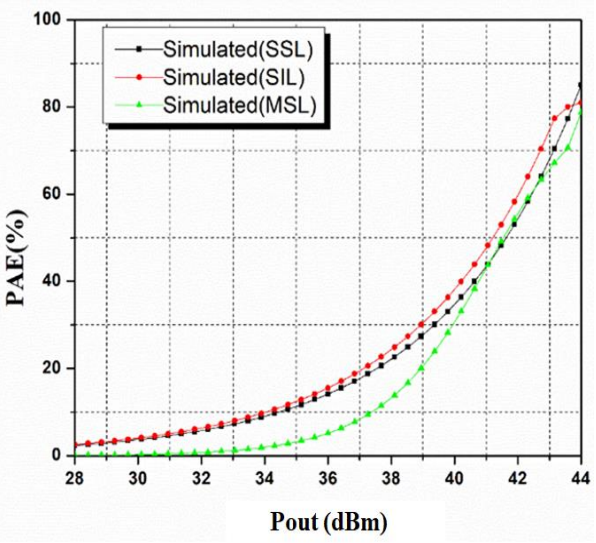


FIGURE 6. PAE (%) Vs Pout (dBm) comparison using SSL, SIL and MSL

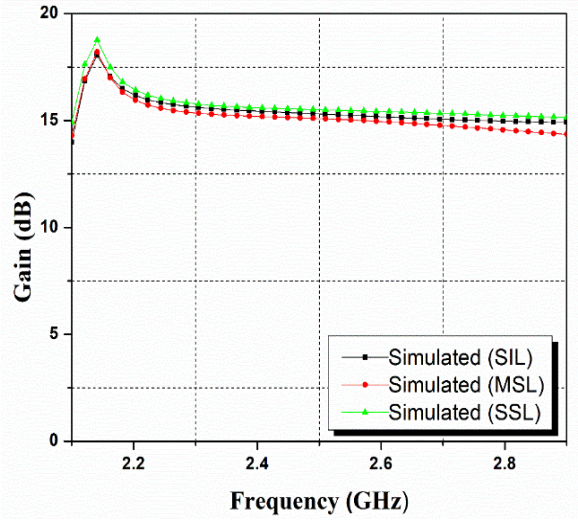


FIGURE 7. Gain (dB) Vs Frequency comparison using SSL, SIL and MSL.

21 February 2024 13:04:41

TABLE 1. Comparison Table of Proposed work with previous work

References	Frequency (GHz)	PAE (%)	Pout (dBm)	Gain (dB)
[17]	2.4GHz	83%	23.65	13.65
[18]	3.1GHz	82%	40	15
[19]	1.5GHz	46.6%	46.9	17.2
[20]	1.8GHz	60%	37	17
This Work	0-3GHz	80-85%	44	15

CONCLUSION

In this paper, a comparison of microstrip line (MSL), suspended shielded line (SSL) and shielded inverted line based matching network is performed for Class-F power amplifier design for smart parking solution using advance design system (ADS) software and CGH00400F GaN HEMT device from CREE. SSL/SIL shows better results than MSL because of low attenuation, small dielectric constant, low propagation loss and low insertion loss. Cascade topology is used for improved gain and efficiency. SSL/SIL shows the better results in comparison to MSL and obtained PAE between 80 to 85%, output power of 44dBm and gain of 15dBm at 0 to 3GHz frequency band.

ACKNOWLEGMENT

Authors would like to thank Department of Electronics and Communications, Delhi Technological University, Delhi, and all its faculty and staff members for their support in research.

REFERENCES

[1] T. Lin, H. Rivano, and F. Le Mouel, "A Survey of Smart Parking Solutions," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 12, pp. 3229–3253, 2017, doi: 10.1109/TITS.2017.2685143

[2] A. Khanna and R. Anand, "IoT based smart parking system," 2016 Int. Conf. Internet Things Appl. IOTA 2016, pp. 266–270, 2016, doi: 10.1109/IOTA.2016.7562735.

[3] T. Perković, P. Šolić, H. Zargariasl, D. Čoko, and J. J. P. C. Rodrigues, "Smart Parking Sensors: State of the Art and Performance Evaluation," *J. Clean. Prod.*, vol. 262, 2020, doi: 10.1016/j.jclepro.2020.121181.

[4] F. Al-Turjman and A. Malekloo, "Smart parking in IoT-enabled cities: A survey," *Sustain. Cities Soc.*, vol. 49, no. May, 2019, doi: 10.1016/j.scs.2019.101608.

[5] S. Y. Zheng, Z. W. Liu, X. Y. Zhang, X. Y. Zhou, and W. S. Chan, "Design of Ultrawideband High-Efficiency Extended Continuous Class-F Power Amplifier," *IEEE Trans. Ind. Electron.*, vol. 65, no. 6, pp. 4661–4669, 2018, doi: 10.1109/TIE.2017.2772163.

[6] H. Zhao and D. Zhang, "Dual-Band Class-F Power Amplifier with Bandwidth Expansion Based on Sector Microstrip Branch for 5G Applications," *IEEE Microw. Wirel. Components Lett.*, vol. 31, no. 7, pp. 877–880, 2021, doi: 10.1109/LMWC.2021.3074411.

[7] A. Pirasteh, S. Roshani, and S. Roshani, "A modified class-F power amplifier with miniaturized harmonic control circuit," *AEU - Int. J. Electron. Commun.*, vol. 97, pp. 202–209, 2018, doi: 10.1016/j.aeue.2018.10.022.

[8] Z. Cheng, X. Xuan, H. Ke, G. Liu, Z. Dong, and S. Gao, "Power Amplifier with Parasitic Compensation Circuit," vol. 2017, no. 2, 2017.

21 February 2024 13:04:41

[9] A. Imtiaz, J. Lees, H. Choi, and L. T. Joshi, “An Integrated Continuous Class-F Mode Power Portable Diagnostic Applications,” vol. 63, no. 10, pp. 3007–3015, 2015.
[10] H. Huang *et al.*, “Design of Multioctave Bandwidth Power Amplifier Based on Resistive Second-Harmonic Impedance Continuous Class-F,” vol. 27, no. 9, pp. 830–832, 2017.
[11] M. Hayati, S. Zarghami, and A. Grebennikov, “Design of a Compact 2.4 GHz Class-F Power Amplifier with High Power Added Efficiency,” *IETE J. Res.*, vol. 0, no. 0, pp. 1–8, 2019, doi: 10.1080/03772063.2019.1644209.
[12] A. M. D. Hbt, S. Muhammad, A. Ali, S. M. R. Hasan, and S. Member, “A 38-GHz Millimeter-Wave Double-Stacked HBT Class-F − 1 High-Gain Power Amplifier in 130-nm SiGe-BiCMOS,” pp. 1–6, 2020.
[13] L. Ma, J. Zhou, and Z. Yu, “Design of a class-f power amplifier with expanding bandwidth,” Asia-Pacific Microw. Conf. Proceedings, APMC, vol. 1, pp. 0–2, 2016, doi: 10.1109/APMC.2015.07411818.
[14] S. K. Dhar *et al.*, “Input-Harmonic-Controlled Broadband Continuous Class-F Power Amplifiers for Sub-6-GHz 5G Applications,” pp. 1–14, 2020.
[15] M. Ekhteraei, M. Hayati, and F. Shama, “High-Efficiency Low Voltage Inverse Class-F Power Amplifier Design Based on Harmonic Control Network Analysis,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 67, no. 3, pp. 806–814, 2020, doi: 10.1109/TCSI.2019.2952932.
[16] Y. Bekri, N. Benabdallah, N. Benhamed, F. T. Bendimerad, and K. Aliane, “Analysis and Design of Shielded Suspended and Inverted Microstrip Lines for Microwave Applications,” *Comput. Sci. Technol. Int. J.*, vol. 2, no. 1, pp. 33–38, 2012, [Online]. Available: http://www.researchpub.org/journal/cstij.
[17] M. Hayati, S. Zarghami, and A. Grebennikov, “Design of a Compact 2.4 GHz Class-F Power Amplifier with High Power Added Efficiency,” *IETE J. Res.*, vol. 0, no. 0, pp. 1–8, 2019, doi: 10.1080/03772063.2019.1644209.
[18] D. Peroulis, “A 3.1-GHz Class-F Power Ampli fi er With 82% Power-Added-Ef fi ciency,” *IEEE Microw. Components Lett.*, vol. 1, no. 8, pp. 3–5, 2013.
[19] H. Wu and G. R. Branner, “High Power Class F GaN HEMT Power Amplifier in L band for Global Positioning Systems Application,” *2018 IEEE 19th Wirel. Microw. Technol. Conf.*, pp. 1–4, 2018, doi: 10.1109/WAMICON.2018.8363920.
[20] T. Sharma, J. S. Roberts, R. Darraji, D. G. Holmes, J. K. Jones, and F. M. Ghannouchi, “Novel Integrated Class F Power Amplifier Design for RF Power Infrastructure Applications,” *IEEE Access*, vol. 6, no. c, pp. 75650–75659, 2018, doi: 10.1109/ACCESS.2018.2881685.

21 February 2024 13:04:41

View publication stats

corrected Proof

Water Science & Technology

© 2024 The Authors

Water Science & Technology Vol 00 No 0, 1 doi: 10.2166/wst.2024.025

Approximation in scour depth around spur dikes using novel hybrid ensemble data-driven model

Balraj Singh* and Vijay K. Minocha

Civil Engineering Department, Delhi Technological University, Delhi 110042, India
*Corresponding author. E-mail: balrajsingh_2k21phdce12@dtu.ac.in

ABSTRACT

The scouring process near spur dikes poses a threat to riverbank stability, making it crucial for river engineering to accurately calculate the maximum scour depth. However, determining the maximum scour depth has been challenging due to the intricacy of scour phenomena surrounding these structures. This research introduces a reliable ensemble data-driven model by hybridizing random tree (RT) using additive regression (AR), bagging (B), and random subspace (RSS) for predicting scour depths around spur dikes. A database of 154 experimental observations was collected from literature, with 103 and 51 observations used for training and testing subsets, respectively. A dimensionless analysis was performed on the collected dataset, selecting four variables as input variables (v/v_s, y/l, l/d_{50}, and Fd_{50}) and d_s/l as response variables. The performance comparison demonstrates that B_AR_RT has a better coefficient of determination (R^2) of 0.9693, root mean square error (RMSE) of 0.1305, and Nash–Sutcliffe efficiency (NSE) of 0.9692. Finally, a comparison of the best hybrid model has been done with previous studies, and sensitivity analysis is performed to determine the most influential parameter for predicting the scour depth around spur dikes.

Key words: ensemble data-driven models, hybrid models, scour depth, spur dikes

HIGHLIGHTS

- Development and inter-comparison of the multi-order novel hybrid ensemble data-driven models using additive regression, bagging, and random subspace.
- The B_AR_RT, a novel hybrid ensemble model, has emerged as the most proficient model in predicting scour depth.
- The sensitivity analysis is performed, and the dimensionless variable, l/d_{50}, is the most influential parameter in predicting scour depth around spur dikes.

1. INTRODUCTION

Scour is the process by which silt, soil, or rock material is eroded or removed from the bed or banks of a river, channel, or coastal region as a result of the movement of water (Yeganeh-Bakhtiary *et al.* 2017, 2020). It happens when the power of the water flow surpasses the soil or sediment's capacity to withstand erosion, resulting in the excavation and deepening of the channel or the destabilization of structures, such as bridges, spur dikes, piers, and embankments. In general, scour may be classified into two types: general scour and local scour. Scour is not limited to a particular building; it can also occur across a larger area (Pandey *et al.* 2018). It involves sediment extraction from the bed or banks of a river or channel due to alterations made to the channel's form. Local scour refers to silt erosion close to the structure, such as bridge piers, abutments, or other obstacles in a river or channel (Singh *et al.* 2022). The presence of the structure leads to an augmentation in flow rates and turbulence. The phenomenon of local scour has the potential to erode the foundation of a building, hence causing issues with its stability. The collapse of over 350 bridges was attributed to local scour around piers and abutments, as shown in a study conducted by the US Federal Highway Administration in 1973 (Richardson & Davis 2001). The collapse of these hydraulic systems demonstrates the criticality of accurately forecasting the levels of erosion around such structures.

Spur dikes are often employed in construction in river engineering and coastal management to control erosion and sediment accumulation. These artificial structures, referred to as groins, are commonly constructed perpendicular to the current of a river or seashore. These constructions protrude from the coastline or riverbank into the ocean (Kothyari *et al.*

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2024.025/1366312/wst2024025.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

corrected Proof

2007; Van Der Wal 2020). They alter the hydraulic conditions and divert the movement of water and sediment, reducing erosion and aiding in the formation or maintenance of stable channel beds. They are constructed using artificial materials like stones, cement, wood, or geotextile tubes. Figure 1 displays the flow and scour pattern schematic representation for a spur dike. The dimensions, intervals, and configuration of spur dikes can differ based on specific project specifications and surrounding ecological factors (Pourshahbaz *et al.* 2022). Spur dikes primarily regulate sediment movement and maintain riverbanks by redirecting water flow toward the middle of the channel. This redirection helps decrease the erosive impact on the banks (Kuhnle *et al.* 2002; Patel *et al.* 2023).

Numerous scholars have studied calculating the scour depth (Karami *et al.* 2014; Özyaman *et al.* 2022). In their study, Pandey *et al.* (2016) examined the phenomenon of spur dike scour, assessed the accuracy of equations used to calculate scour depth, and developed an empirical equation for determining the maximum scour depth. In their study, Kothyari et al. (2007) examined the effects of single and multiple spur dikes on the changes in scour depth over time. Subsequently, the researchers formulated an equation to forecast the fluctuation in scour depth over time for piers, abutments, and spur dikes. Zhang & Nakagawa (2009) examined the flow pattern and properties of bed deformation in the vicinity of spur dikes. In their study, Nath & Misra (2017) examined how the length of spur dikes affects the scouring process. The procedure of experimentally determining the scour depth surrounding spur dikes is both time-consuming and difficult. However, this may be substituted by using data-driven models.

In the past decade, data-driven models have been extensively used to address a wide range of challenging hydraulics and water resource engineering problems (Arora *et al.* 2019; Singh 2020; Bhoria *et al.* 2021; Sepahvand *et al.* 2021; Donnelly *et al.* 2022, 2023, 2024; Yeganeh-Bakhtiary *et al.* 2022, 2023; Kumar *et al.* 2023a, 2023b). Recent research has emphasized using data-driven models to predict the maximal scour depth near various hydraulic structures (Khosravi *et al.* 2021; Pandey *et al.* 2021; Kumar & Afzal 2023; Marulasiddappa *et al.* 2023). Kumar *et al.* (2023a, 2023b) extensively studied the scour phenomena using many empirical and data-driven approaches to estimate the scour depth. Out of two ensemble methods (Bagging Regressor, AdaBoost Regressor) and one standalone machine learning method (Support Vector Regression), both the ensemble methods provide superior predictions compared to the standalone variant and empirical equations. Pandey *et al.* (2021) used three innovative ensemble-based data intelligence paradigms, namely categorical boosting (CatBoost) in conjunction with extra tree regression (ETR) and K-nearest neighbor (KNN), for predicting the scour depth. They found that the CatBoost model provided better accuracy than other machine learning models. Chou & Nguyen (2022) proposed a novel artificial intelligence-based approach by integrating a forensic-based investigation (FBI) algorithm, least squares SVR, and radial basis function neural network to assist civil engineers in estimating scour depth. The proposed approach performs more precisely than other artificial intelligence methods. Aamir & Ahmad (2019) deployed the adaptive neuro-fuzzy interference system (ANFIS) and artificial neural network (ANN) to compute the maximum scour depth downstream of a rigid apron under wall jets. They discovered that the ANFIS and ANN models provided superior results compared to previously proposed regression and empirical equations. Tripathi & Pandey (2023) utilized gene expression programming and Gaussian process regression to predict the scour depth. They concluded that gene expression programming (GEP) is the alternative model that can be used to predict the scour depth around spur dikes.

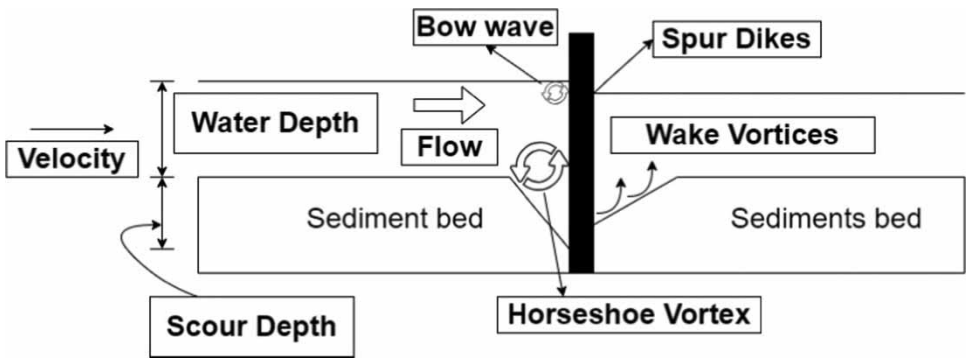


Figure 1 | Diagram of the scour and flow pattern around a spur dike.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2024.025/1366312/wst2024025.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

corrected Proof

Although several data-driven models have been employed in past research, hybrid data-driven models still need to be applied for modeling the scour depth around spur dikes (Minocha & Singh 2023). Therefore, the following objectives were formulated for the present study to spotlight the gap: (1) development of hybrid data-driven models for predicting scour depth for spur dikes for the first time. This can be achieved by creating the appropriate hybrid data-driven model using bagging, additive regression (AR), and random subspace (Elbeltagi *et al.* 2023; Pande *et al.* 2023); (2) determination of the top-performing data-driven models using three different performance evaluation metrics (i.e., R^2, coefficient of determination (COD) and Nash–Sutcliffe efficiency (NSE)), two mathematical diagrams (box plots, and Taylor diagrams for both the training and testing datasets; and (3) verification of the adaptability and versatility of the proposed data-driven model by comparing its performance with models used in the previous study.

2. MATERIALS AND METHOD

This research consists of several processes, such as data collection, dimensionless analysis, dataset, data-driven standalone or conventional models, hybrid models, and performance evaluation of data-driven models. The workflow of the study is displayed in Figure 2. The details of these processes are described in the following subsections:

2.1. Data collection and dimensionless analysis

The dataset of Ezzeldin *et al.* (2007), Nasrollahi *et al.* (2008), Pandey *et al.* (2016), and Zaghloul (1983) has been used in the study which consists of the length of spur dikes (l) in meters, flow velocity (v) in meters per second, threshold velocity (v_s) meters per second, flow depth (y) in meters, median particle size diameters (d_{50}) in meters, the width of the flume (B) in meters, and scour depth (d_s) in meters. Table 1 illustrates the range and statistical description of parameters obtained (l, v, y, d_{50}, Fd_{50}, B, and Reynolds number) from the abovementioned studies for individual datasets. In the overall dataset, v varied from 0.1590 to 0.6700, v_s varied from 0.1825 to 0.70590, y varied from 0.0380 to 0.6000, l varied from 0.0500 to 4.7500, d_{50} varied from 0.0003 to 0.0031, and d_s varied from 4.7000 to 74.7000. Additionally, the dimensionless analysis is done by Pandey *et al.* (2021), and the dimensionless equation becomes:

$$\frac{d_s}{l} = f\left(\frac{v}{v_s},\ \frac{y}{l},\ \frac{l}{d_{50}},\ Fd_{50}\right) \tag{1}$$

The collected raw data were processed as the independent dimensionless variables (v/v_s, y/l, l/d_{50}, and Fd_{50}) as the input and the maximum scour depth (d_s/l) as the response variable. The total dataset (154 observations) was statically analyzed and divided into two groups. The larger group included 103 data used to create models, known as the training subset, while the other group included 51 data used to validate models (testing subset). Table 2 provides the statistical description of the input and output variables according to the training and testing subsets. The statistical parameters used to describe data are minimum, maximum, mean, and standard deviation.

2.2. Ensemble data-driven models

The data-driven models are one of the modern and advanced methods used in water resource and hydraulics engineering (Sharafati *et al.* 2020; Khosravi *et al.* 2022; Mamun *et al.* 2022; Nourani *et al.* 2022; Niazkar & Zakwan 2023). Ensemble data-driven models refer to the approach of combining several types of models to improve efficiency and robustness. These models can mitigate overfitting, a condition characterized by a model exhibiting strong performance on training data but poor performance on testing data. By integrating many models with different sources of error, these models can improve the precision and reliability of predictions (Rodriguez-Lozano *et al.* 2020; Jung & Schindler 2023). Weka software has been used in this study for various computing and optimization of data-driven models (Kumar *et al.* 2022). The descriptions of these data-driven models are as follows.

2.2.1. Random tree

The random tree (RT) algorithm addresses classification and regression problems. Aldous (1993) initially established the RT algorithm. It generates a random data collection process for constructing a decision tree. The optimal split of all variables is used to divide each node in a typical tree. Each node is divided using the highest randomly selected subset of predictors in a random forest (RF). The RT node uses bootstrap sampling with replacement to generate sample data. They are using the sample data to develop a tree model. As the tree grows, RT will not sample the data again (Pham *et al.* 2022). Instead, it

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2024.025/1366312/wst2024025.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

corrected Proof

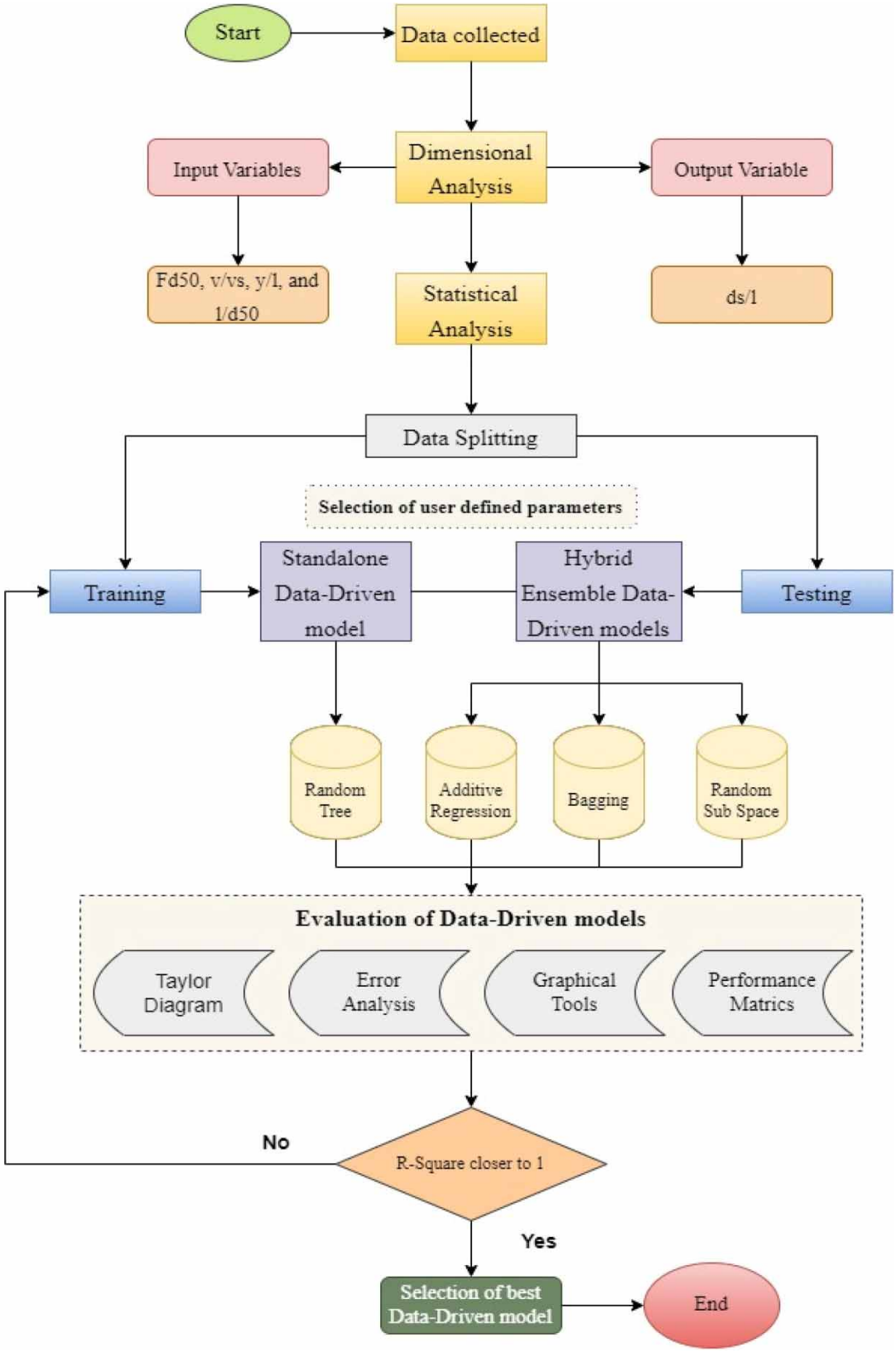


Figure 2 | Flow diagram for predicting the scour depth around spur dikes.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2024.025/1366312/wst2024025.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

corrected Proof

Table 1 | Range of the parameters obtained from previous studies

Studies	l	v	y	d_{50}	Fd_{50}	B	Reynolds number
Zaghloul (1983)	0.045–0.23	0.32–0.43	0.135–0.30	4.49	1.69	0.46	12,700–135,500
Ezzeldin *et al.* (2007)	0.04–0.1	0.195–0.31	0.065–0.13	0.60	1.98–3.15	0.40	18,600–36,100
Nasrollahi *et al.* (2008)	0.25–0.75	0.3125	0.128	1.3	1.467–2.15	2.0	40,000
Pandey *et al.* (2016)	0.06–0.15	0.21–0.43	0.05–0.53	0.8–1.02	1.78–4.07	1.0	12,700–135,500

Table 2 | Statistical description of the dimensionless data

Statistical parameters	Minimum	Maximum	Mean	Standard deviation
Training				
Fd_{50}	1.4670	4.0695	3.0497	0.4590
v/v_s	0.5214	3.5121	0.9894	0.4212
y/l	0.0126	4.1667	2.0229	1.1585
l/d_{50}	19.3548	5277.7778	342.8931	790.0412
d_s/l	0.0057	4.0456	1.7319	0.7737
Testing				
Fd_{50}	1.7776	3.9616	3.0485	0.5146
v/v_s	0.5105	2.5257	0.9076	0.2678
y/l	0.0343	4.1667	1.6760	1.1170
l/d_{50}	25.8065	1944.4444	279.1039	318.0887
d_s/l	0.3041	3.2293	1.5992	0.7519

randomly selects a subset of predictors and uses the strongest one to divide a tree node. The process is iterated while partitioning each tree node and constructing a tree in an RF is based on this fundamental premise. RT models are reliable when dealing with large-scale datasets and numerous variables. By employing tagging and field sampling techniques, the susceptibility to overfitting is significantly reduced, increasing the likelihood of replicating test results with new data.

2.2.2. Additive regression

The AR algorithm, initially proposed by Breiman & Friedman (1985), is a nonparametric method that plays a vital role in the alternating conditional-expectations algorithm. AR employs a one-dimensional smoother to create a restricted collection of nonparametric regression models that offer greater flexibility than linear models and are easier to comprehend than generic ones. To summarize, the AR algorithm may be defined as follows.

Consider a dataset $Z = \{(Y_h, X_h), h = 1, 2, \ldots., N\}$, here Y_h is the output variable and X_h is the input variable. The AR model will take the form:

$$E[Y|X_1, \ldots., X_k] = \beta_0 + \sum_{j=1}^{k} f_j(X_{ij}) \tag{2}$$

The dataset is fitted using unknown smooth functions $f_j(x_{ij})$.

2.2.3. Bagging

Slight modifications in the dataset can lead to instability in data-driven models. In 1996, Breiman proposed a technique called 'bagging' to tackle the problem of model instability and improve overall performance. Each predictor is provided with a subset of the primary dataset in this approach. In other words, each predictor only has access to a subset of the data and must

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2024.025/1366312/wst2024025.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

corrected Proof

construct its model using only that subset. Nguyen & Bui (2019) have demonstrated the use of this method in several studies to improve prediction accuracy.

2.2.4. Random subspace

The random subspace (RSS) strategy utilizes bootstrapping and aggregation to develop a predictive model similar to the bagging method. Unlike bagging, RSS utilizes bootstraps in the feature space instead of the training samples (Tao *et al.* 2006). Like RF and B, RSS can process datasets for regression and classification applications. The system consists of four main components: the training dataset, labeled as x; the number of subspaces, represented by L; the classifier or regressor, marked as w; and the number of features, indicated by Ds (Skurichina & Duin 2002). The number of subgroups in RSS is randomly generated based on the properties of the data structure. Subsequently, they are stored in the subspace L. In the second phase, separate base regressors/classifiers are trained on each subset to produce unique regressors/classifiers. Afterward, the components mentioned above are combined to create an ensemble regressor/classifier called E. For further in-depth knowledge on RSS, refer to the publication by Kuncheva & Plumpton (2010).

2.2.5. Development of hybrid ensemble data-driven models

This research has developed hybrid models to predict the scour depths around the spur dikes. The standalone data-driven model (RT) is hybridized using AR, bagging, and random sub-surface (RSS). With the combination of the standalone model (RT) and AR, B, and RSS, seven hybrid models were developed. These seven hybrid data-driven models were AR_RT, B_RT, RSS_RT, RSS_AR_RT, RSS_B_RT, B_AR_RT, and RSS_B_AR_RT. Finally, the performance of the hybrid data-driven models was compared with the standalone model (RT) to predict scour depths around spur dikes.

2.3. Performance evaluation

In the present research, four performance evaluation indicators have been used to measure the performance of different data-driven models. These performance indicators are root-mean-square error (RMSE), coefficient of determination (R^2), and Nash–Sutcliffe efficiency (NSE). The mathematical formulation of the indicators is as follows.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{k=1}^{t} (\mathrm{Scour}_{\mathrm{Actual},\,k} - \mathrm{Scour}_{\mathrm{Predicted},k})^2}{t}} \tag{3}$$

$$\mathrm{COD} = \frac{\mathrm{TSS} - \mathrm{RSS}}{\mathrm{TSS}} \tag{4}$$

$$\mathrm{NSE} = 1 - \frac{\sum_{k=1}^{t} (\mathrm{Scour}_{\mathrm{Actual},\,k} - \mathrm{Scour}_{\mathrm{Predicted},k})^2}{\sum_{k=1}^{t} (\mathrm{Scour}_{\mathrm{Actual},k} - \overline{\mathrm{Scour}}_{\mathrm{Actual}})^2} \tag{5}$$

where $\mathrm{TSS} = \sum_{k=1}^{t} (\mathrm{Scour}_{\mathrm{Actual},k} - \overline{\mathrm{Scour}}_{\mathrm{Actual}})^2$, $\mathrm{RSS} = \sum_{k=1}^{t} (\mathrm{Scour}_{\mathrm{Predicted},k} - \mathrm{Scour}_{\mathrm{Actual},\,k})^2$, $\mathrm{Scour}_{\mathrm{Actual}}$ denotes the actual values of scour depths, $\mathrm{Scour}_{\mathrm{Predicted}}$ denotes the predicted values of scour depths by data-driven models, $\overline{\mathrm{Scour}}_{\mathrm{Actual}}$ is the mean of the actual values of scour depths, $\overline{\mathrm{Scour}}_{\mathrm{Predicted}}$ is the mean of the predicted values of scour depths, and t denotes the number of observations. Furthermore, the utilization of the Taylor diagram (Taylor 2001) and box plot (Williamson *et al.* 1989) was employed to facilitate the visual evaluation of the models. The Taylor diagram comprehensively represents each model's correlations and standard deviations. Greater predictive capacity is achieved when the predicted value of the Coefficient of Correlation (CoC) and standard deviation approach the observed value. The box plot comprehensively evaluated the models' capacities using extreme values, medians, and first- and third-quartile predictions.

2.4. User-defined parameters

A total of eight data-driven models (one standalone model, RT, and seven hybrid models; AR_RT, B_RT, RSS_RT, RSS_AR_RT, RSS_B_RT, B_AR_RT, RSS_B_AR_RT) are used in this study to predict the scour depth around spur dikes. The performance of the data-driven models strictly depends upon the user-defined parameters (UDFs). Several combinations of UDFs are used; on these sets, the performance of the data-driven models is checked. The set that gives the best results of

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2024.025/1366312/wst2024025.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

corrected Proof

data-driven models is chosen. These chosen values of UDFs for different used data-driven models for the prediction of scour depth in this study are summarized in Table 3, along with the input and output variables.

3. RESULTS AND DISCUSSIONS

The performance evaluation is the measure that checks the performance of the data-driven models. This study uses three statistical parameters (Equations (3)–(5)). The performance evaluation outcomes for various data-driven models are shown in Table 4 for the training and testing subsets. Evidently, the B_AR_RT model predicted greater efficiency in testing phase ($R^2 = 0.9702$, RMSE = 0.1305, and NSE = 0.9692) followed by RSS_RT; B_RT; AR_RT; RSS_B_RT; RSS_B_AR_RT; and RT. Furthermore, the difference of R^2, RMSE, and NSE between the actual values and predicted values at training and testing subsets for B_AR_RT is considerably less than other hybrid and standalone models. So, it is concluded from Table 4 that B_AR_RT is the most efficient data-driven model for the prediction of scour depth. These conclusions also suggest that first-order hybridization increases the efficiency of the data-driven model. So, RSS_RT, B_RT, and AR_RT give better results than the standalone model (RT). Second-order hybridization (B_AR_RT) also increases the efficiency of the data-driven models except RSS_AR_RT and RSS_B_RT and has the highest accuracy among all data-driven models. However, third-order hybridization decreases the performance, but its performance is better than the standalone model. Thus, the B_AR_RT data-driven model gives the best result for predicting scour depth and has supremacy among all competing data-driven models.

To demonstrate the performance at training and testing of such data-driven models, several plotting techniques such as scatter plot, variation plot, box plot, and Taylor diagram have been used in the literature (Malik *et al.* 2022; Jaiswal *et al.* 2023; Salmasi *et al.* 2023). In this research, the authors have used these plotting techniques to validate the results of the performance indicators. The plots of this study for the prediction of scour depths around spur dikes are as follows:

A. The scattered plots of scour depths with all data-driven models for training and testing subsets are plotted in Figure 3(a) and 3(b), respectively. It is plotted between the actual and predicted values of dimensional output variables (d_s/l), and the

Table 3 | UDFs of data-driven models

Data-driven models	UDFs	Input and output variables
Random tree	Batch_size-100, max_depth = 0, Min_Number = 1, Min_variance_proportion = 0.001, and random_seed = 1	v/v_s, y/l, l/d_{50}, Fd_{50}, and d_s/l
Bagging	Batch_size – 100, bag_Size_percent = 100	
AR	Batch size – 100,	
RSS	Batch_size – 100	

Training and testing percentages = 67 and 33%, respectively.

Table 4 | Results of performance evaluation for different data-driven models

	R^2	RMSE	NSE	R^2	RMSE	NSE
Data-driven models	Training			Testing		
RT	0.9998	0.0106	0.9998	0.9081	0.2308	0.8039
AR_RT	1	0.0002	1	0.9502	0.1699	0.948
B_RT	0.9866	0.0927	0.9855	0.9609	0.1506	0.9591
RSS_RT	0.9944	0.0605	0.9938	0.9699	0.1321	0.9685
RSS_AR_RT	0.9504	0.1887	0.94	0.8967	0.2463	0.8905
RSS_B_RT	0.937	0.2253	0.9143	0.9192	0.2363	0.8993
B_AR_RT	0.9873	0.0899	0.9864	0.9702	0.1305	0.9692
RSS_B_AR_RT	0.937	0.2253	0.9144	0.9129	0.2383	0.8947

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2024.025/1366312/wst2024025.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

corrected Proof

diagonal line is the best fit. The model in which all the points lie on best-fit lines is the best. Figure 3(a) and 3(b) also suggested that B_AR_RT is the best data-driven model, while RSS_B_AR_RT gives the worst result among all data-driven models in predicting scour depths. The results of the scattered plots also follow the same hierarchy, followed by the various performance indicators.

The variation plots of various data-driven models are depicted in Figure 4 for training and testing subsets. As the name indicates, the variations plot shows the visual interpretation of the variation among the actual and predicted values of d_s/l. Figure 5 suggests a minimum difference between actual and predicted values of d_s/l for B_AR_RT data-driven models, followed by RSS_RT, B_RT, AR_RT, RSS_B_RT, RSS_B_RT_RT, RT, and RSS_AR_RT. The variation plot (Figure 4) suggests the same trend as in Table 4 and Figure 3(a).

B. The distributions of the actual and predicted values of d_s/l in the form of an open box plot for all data-driven models are plotted in Figure 5(a) and 5(b) to illustrate the efficacy of the data-driven models. The box-and-whisker plots display the measures of the variables' central tendency, variability, skewness, and extreme values. These graphs visually represent the local discrepancies between predicted outputs and actual data based on four components: (1) Each box contains a central solid line, representing the median value indicating the central tendency or location. (2) The box itself represents the variability around this central tendency, with the edges of the box representing the 25th and 75th percentiles. (3) Whiskers extend from the box to display the range of the data. (4) Values that fall outside the length of the whiskers are considered outliers and are represented by a black diamond (◆) symbol. These outliers have a value that is more than 1.5 times the interquartile range (IQR = Q3–Q1) away from the top or bottom of the box. Upon examination of Figure 5, the descriptive statistics indicate that the lowest, maximum, 25th quartile value (Q1), 75th quartile value (Q3), and the median value (Q2)

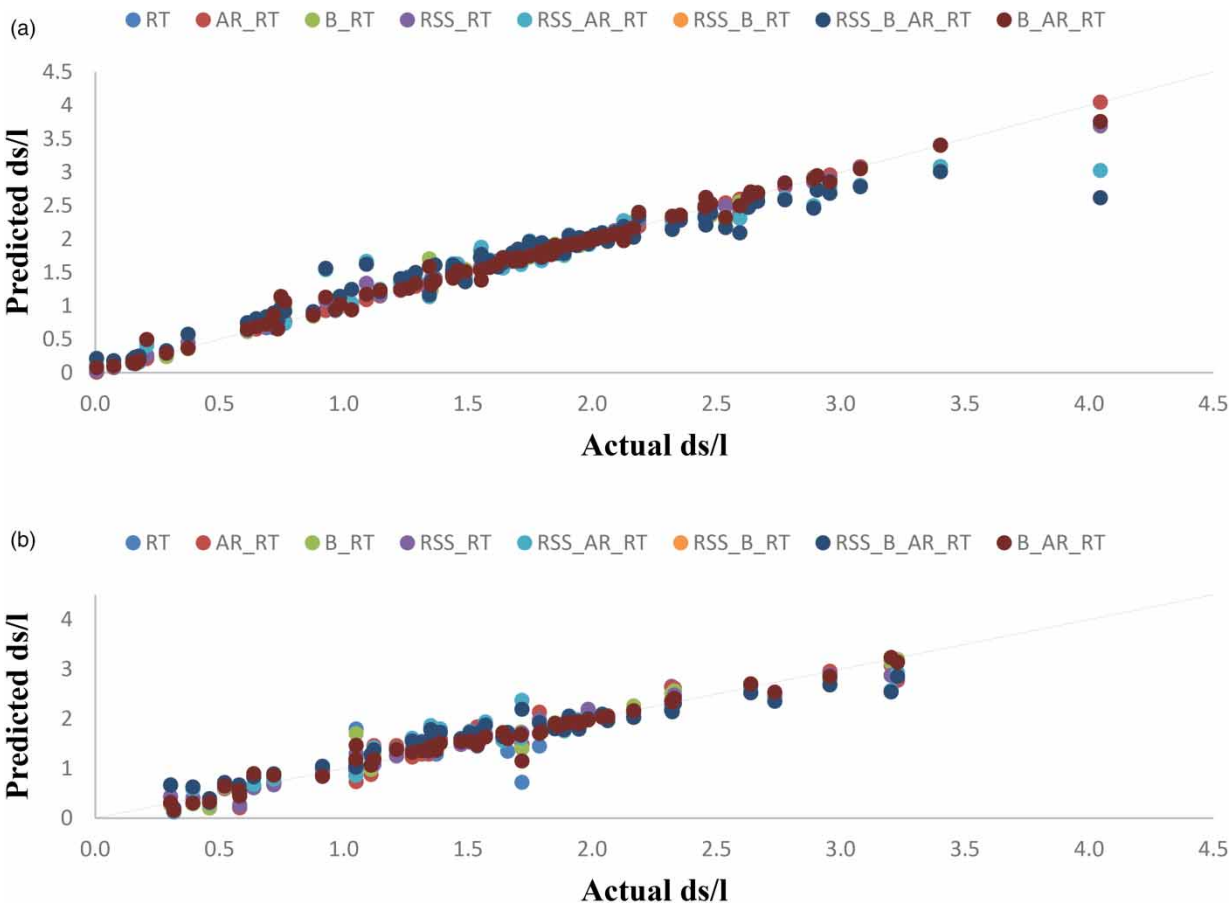


Figure 3 | Scattered plot between the actual and predicted values of d_s/l for (a) training subsets and (b) testing subsets.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2024.025/1366312/wst2024025.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

corrected Proof

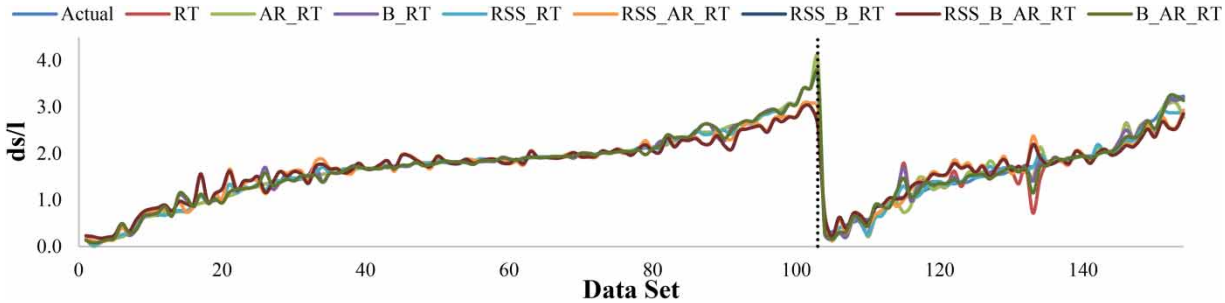


Figure 4 | Variation plots of various data-driven models in the prediction of d_s/l.

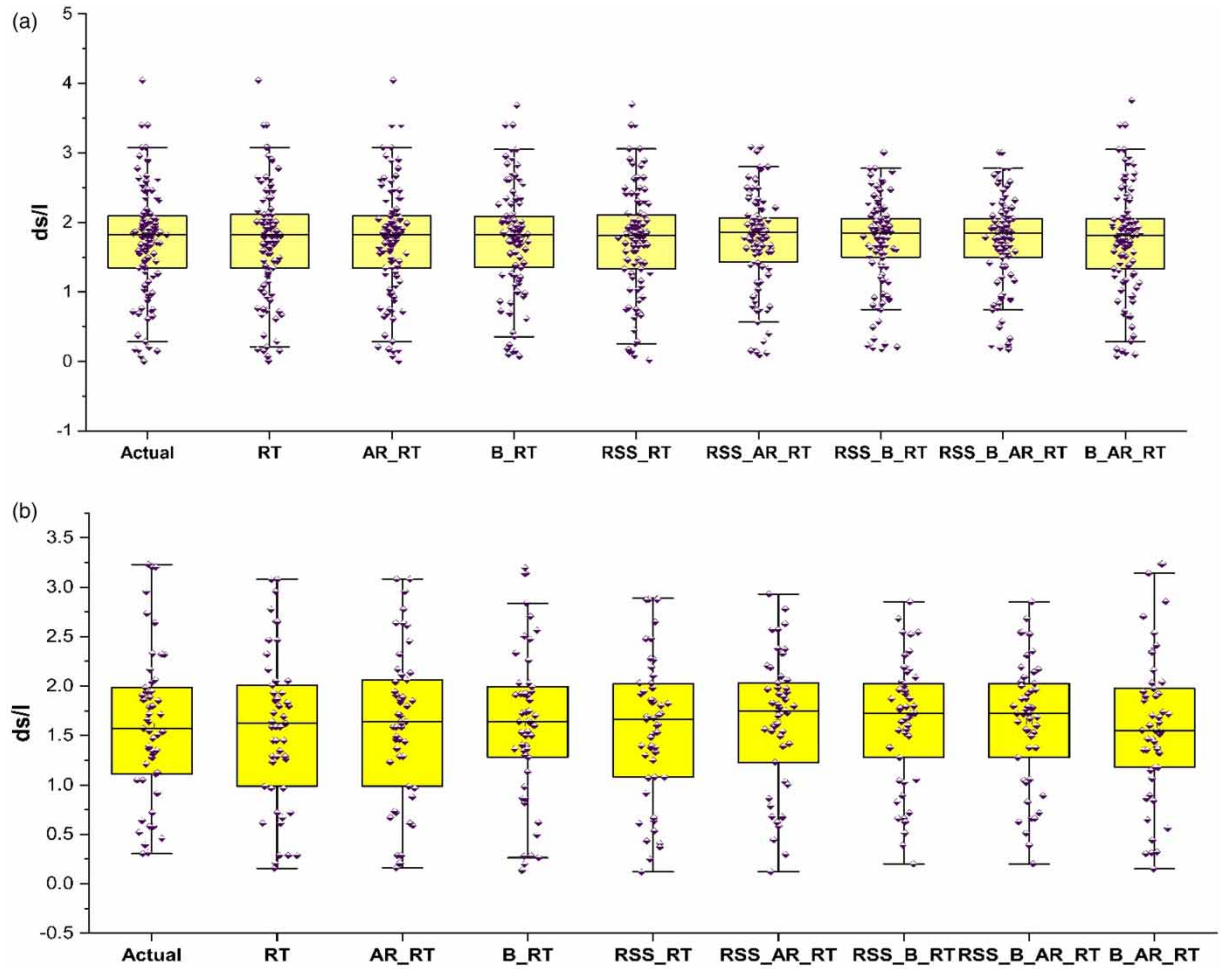


Figure 5 | Box plot of different data-driven models predicting d_s/l for (a) training and (b) testing subsets.

of B_AR_RT and actual values are quite similar. When comparing the observed data with the results of the B_AR_RT-based data-driven model, the widths of the top and lower boxes are almost identical, as shown in Figure 5(a) and 5(b). Also, the box plots suggest that the accuracy order of the different data-driven models is: B_AR_RT > RSS_RT > B_RT > AR_RT > RSS_B_RT > RSS_B_RT_RT > RT > RSS_AR_RT for testing subsets.

C. The Taylor diagrams for the data-driven model are displayed in Figure 6. The Taylor diagram is a widely used graphical tool for comparing data-driven models fairly. The Taylor diagram quantifies the level of agreement between observed and

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2024.025/1366312/wst2024025.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

corrected Proof

predicted values based on three statistical measures: RMSE, standard deviation, and correlation coefficient (CC). Among the models depicted in this mathematical diagram, the one that exhibits superior performance is the model positioned closer to the actual value point, represented by a hollow black circle. The Taylor diagram findings demonstrated that the B_RT_RT-based model exhibited superior accuracy to the other data-driven models for predicting the scour depth around spur dikes.

For the measurement of the scour depth around spur dikes, Tripathi & Pandey (2023), Pandey *et al.* (2021), and Karami *et al.* (2012) have used data-driven models to predict scour depth around spur dikes. So, a comparison of the best model of this study (B_AR_RT) is done with previous studies (Karami *et al.* 2012; Pandey *et al.* 2021; Tripathi & Pandey 2023). The results were compared based on the performance evaluation (R^2, RMSE, and R) and tabulated in Table 5. The results suggest that the best hybrid model of this study (B_AR_RT) also got the first rank based on the performance evaluation, while stacked (boosting regression tree (BRT)) of Pandey *et al.* (2021) got the second rank and multi gene genetic programming (MGGP) of Pandey *et al.* (2021) got the last rank. ANN_FFBP (feed-forward backpropagation) and ANN_RBF (Radial Basis Function) of Karami *et al.* (2012) got sixth and seventh rank, respectively. So, the proposed novel hybrid ensemble model is highly accurate and more robust in the prediction of the scour depth around spur dikes not only in the competing hybrid and standalone models but also in various developed models by Tripathi & Pandey (2023), Pandey *et al.* (2021) and Karami *et al.* (2012).

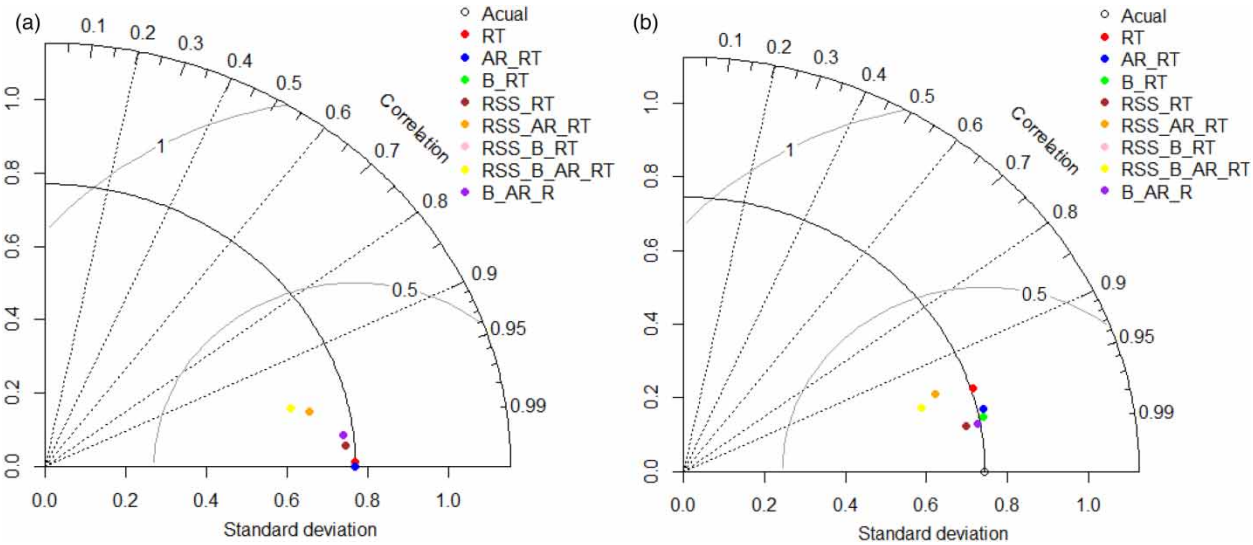


Figure 6 | Taylor diagram of different data-driven models in the prediction of d_s/l for (a) training and (b) testing subsets.

Table 5 | Comparison of results with previous studies

Studies	Model	Performance evaluation			Rank
		R^2	RMSE	R	
This study	B_AR_RT	0.9702	0.1305	0.9850	1
Tripathi & Pandey (2023)	GEP	0.9400	–	0.9695	3
Pandey *et al.* (2021)	MARS	0.9332	0.2116	0.9660	5
	MGGP	0.9116	0.2385	0.9548	8
	KELM	0.9384	0.1958	0.9687	4
	Stacked (BGT)	0.9490	0.1831	0.9742	3
	Stacked (BRT)	0.9577	0.1654	0.9786	2
Karami *et al.* (2012)	ANN_FFBP	0.93	0.014	0.9644	6
	ANN_RBF	0.92	0.015	0.9592	7

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2024.025/1366312/wst2024025.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

corrected Proof

Table 6 | Sensitivity analysis study utilizing B_AR_RT

Input combinations	Parameter to be removed	Training		Testing	
		R^2	RMSE	R^2	RMSE
v/v_s, y/l, l/d_{50}, Fd_{50}		0.9872	0.0899	0.9694	0.1305
y/l, l/d_{50}, Fd_{50}	v/v_s	0.9845	0.1004	0.9686	0.1326
v/v_s, l/d_{50}, Fd_{50}	y/l	0.9902	0.0824	0.9649	0.1472
v/v_s, y/l, Fd_{50}	l/d_{50}	0.9730	0.1343	0.8917	0.2465
v/v_s, y/l, l/d_{50}	Fd_{50}	0.9892	0.0856	0.9779	0.1132

This study introduced seven hybrid data-driven models in predicting the spur dikes, and the performance of the hybrid models is compared with the standalone model (RT). The results showed that all models except RSS_AR_RT, RSS_B_AR_RT, and RSS_B_RT could predict scour depth satisfactorily. However, based on several performance evaluation indicators and plots, B_AR_RT is the most efficient data-driven model for the prediction of scour depth, followed by RSS_RT, B_RT, AR_RT, RSS_B_RT, RSS_B_RT_RT, RT, and RSS_AR_RT. It is also found that hybridization of the data-driven models improves prediction efficiency. The efficiency increases in the second-order hybridization in case the hybridization of RT is done with AR, followed by B. However, it decreases in the case of AR, followed by RSS, and B, followed by RSS. Third-order hybridization also decreases the efficiency of the data-driven models. Also, the developed best hybrid model (B_AR_RT) outperforms the previous studies. Thus, the B_AR_RT data-driven model, the second-order hybridization model, best predicts scour depth in testing subsets and has supremacy among all hybrid data-driven models.

3.1. Sensitivity analysis

The sensitivity of the input variable to the output was analyzed using the best-performed model (B_AR_RT). Various training datasets were generated by eliminating one-by-one input variables, and results were reported in evaluating metrics R^2 and RMSE. Findings from Table 6 suggest that (l/d_{50}) plays a crucial role in predicting scour depth around spur dikes with respect to other input variables.

4. CONCLUSION

Developing and introducing dependable hybrid ensemble data-driven models for predicting the scour depths around the spur dike is necessary, which can save time and help hydraulics engineers. This research uses one standalone data-driven and seven hybrid models to predict the scour depths around spur dikes. These seven-ensemble data-driven models are made by hybridization of RT by AR, B, and RSS. The performance and accuracy of all data-driven models have been compared and analyzed using RMSE, R^2, and NSE. Comparing performance and accuracy shows that B_AR_RT (with $R^2 = 0.9693$, RMSE = 0.1305, and NSE = 0.9692) has predicted the scour depths with superior performance in this study. The graphical tools (scattered plot, variation plot, box plot, and Taylor diagram) also represent the supremacy of B_AR_RT in predicting the scour depths around spur dikes. The order of the different data-driven models is B_AR_RT > RSS_RT > B_RT > AR_RT > RSS_B_RT > RSS_B_RT_RT > RT > RSS_AR_RT. It is concluded from the research that hybridization increases the accuracy level of the data-driven models. The accuracy level again increases except RSS_AR_RT and RSS_B_RT when double hybridization is done but decreases when more than two hybridizations occur. Furthermore, the results of the best models are compared with the previous studies, and it was found that the best hybrid model outperforms not only the competing model but also various models published in the previous studies. The result of the proposed model is compared with the various models proposed by previous studies, and the proposed model emerges as the most robust model. Finally, the most influential parameters are determined using sensitivity analysis, and l/d_{50} is the parameter that affects the output the most. Thus, B_AR_RT is the most accurate hybrid data-driven model, which can predict the scour depth very precisely, and it will save time for carrying out the arrangement of physical model fabrications and experimentations.

The proposed data-driven model, utilizing hybrid ensemble data-driven models, has several benefits compared to individual models, including enhanced accuracy and performance, particularly for intricate and noisy issues. In addition, they can mitigate the likelihood of overfitting and underfitting by effectively managing the tradeoff between bias and variance and

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2024.025/1366312/wst2024025.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

corrected Proof

employing diverse subsets and characteristics of the data. However, these ensemble data-driven models have several limitations, including their high computational cost and time requirements. This might augment the intricacy and memory demands of the system. Therefore, to avoid these difficulties, it is advisable to further investigate the behavior of the scour depth around spur dikes by employing advanced hybrid soft computing models such as adaptive neuro-fuzzy inference system (ANFIS) combined with genetic algorithm (GA), ANFIS combined with particle swarm optimization (PSO), SVR combined with GA, SVR combined with PSO, SVR combined with firefly algorithm (FFA), and other similar approaches.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Aamir, M. & Ahmad, Z. 2019 Estimation of maximum scour depth downstream of an apron under submerged wall jets. *Journal of Hydroinformatics* **21** (4), 523–540.

Aldous, D. 1993 The continuum random tree III. *The Annals of Probability* **21**, 248–289.

Arora, S., Singh, B. & Bhardwaj, B. 2019 Strength performance of recycled aggregate concretes containing mineral admixtures and their performance prediction through various modeling techniques. *Journal of Building Engineering* **24**, 100741.

Bhoria, S., Sihag, P., Singh, B., Ebtehaj, I. & Bonakdari, H. 2021 Evaluating Parshall flume aeration with experimental observations and advanced soft computing techniques. *Neural Computing and Applications* **33**, 17257–17271.

Breiman, L. & Friedman, J. H. 1985 Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association* **80** (391), 580–598.

Chou, J. S. & Nguyen, N. M. 2022 Scour depth prediction at bridge piers using metaheuristics-optimized stacking system. *Automation in Construction* **140**, 104297.

Donnelly, J., Abolfathi, S., Pearson, J., Chatrabgoun, O. & Daneshkhah, A. 2022 Gaussian process emulation of spatio-temporal outputs of a 2D inland flood model. *Water Research* **225**, 119100.

Donnelly, J., Abolfathi, S. & Daneshkhah, A. 2023 A physics-informed neural network surrogate model for tidal simulations. *ECCOMAS Proceedia* 836–844. doi:10.7712/120223.10379.19908.

Donnelly, J., Daneshkhah, A. & Abolfathi, S. 2024 Forecasting global climate drivers using Gaussian processes and convolutional autoencoders. *Engineering Applications of Artificial Intelligence* **128**, 107536.

Elbeltagi, A., Al-Mukhtar, M., Kushwaha, N. L., Al-Ansari, N. & Vishwakarma, D. K. 2023 Forecasting monthly pan evaporation using hybrid additive regression and data-driven models in a semi-arid environment. *Applied Water Science* **13** (2), 42.

Ezzeldin, M. M., Saafan, T. O., Rage, O. F., Nejm, I. N. & Assistant, T. 2007 Local scour around spur dikes. In: *Proceedings of the 11th International Conference on Water Technology*, London: International Water Technology Association.

Jaiswal, A., Goel, A. & Sihag, P. 2023 Approximation of aeration efficiency at sharp-crested weirs using metaheuristic regression approaches. *Journal of Hydroinformatics* **25** (3), 1084–1102.

Jung, C. & Schindler, D. 2023 Comprehensive validation of 68 wind speed models highlights the benefits of ensemble approaches. *Energy Conversion and Management* **286**, 117012.

Karami, H., Ardeshir, A., Saneie, M. & Salamatian, S. A. 2012 Prediction of time variation of scour depth around spur dikes using neural networks. *Journal of Hydroinformatics* **14** (1), 180–191.

Karami, H., Basser, H., Ardeshir, A. & Hosseini, S. H. 2014 Verification of numerical study of scour around spur dikes using experimental data. *Water and Environment Journal* **28** (1), 124–134.

Khosravi, K., Khozani, Z. S. & Mao, L. 2021 A comparison between advanced hybrid machine learning algorithms and empirical equations applied to abutment scour depth prediction. *Journal of Hydrology* **596**, 126100.

Khosravi, K., Golkarian, A. & Tiefenbacher, J. P. 2022 Using optimized deep learning to predict daily streamflow: A comparison to common machine learning algorithms. *Water Resources Management* **36** (2), 699–716.

Kothyari, U. C., Hager, W. H. & Oliveto, G. 2007 Generalized approach for clear-water scour at bridge foundation elements. *Journal of Hydraulic Engineering* **133** (11), 1229–1240.

Kuhnle, R. A., Alonso, C. V. & Shields Jr, F. D. 2002 Local scour associated with angled spur dikes. *Journal of Hydraulic Engineering* **128** (12), 1087–1093.

Kumar, L. & Afzal, M. S. 2023 Estimating pier scour depth under combined waves and current using boosting machine-learning models. *Acta Geophysica* 1–17. doi:10.1007/s11600-023-01089-2.

Kumar, M., Elbeltagi, A., Pande, C. B., Ahmed, A. N., Chow, M. F., Pham, Q. B., Kumari, A. & Kumar, D. 2022 Applications of data-driven models for daily discharge estimation based on different input combinations. *Water Resources Management* **36** (7), 2201–2221.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2024.025/1366312/wst2024025.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

corrected Proof

Kumar, S., Goyal, M. K., Deshpande, V. & Agarwal, M. 2023a Estimation of time-dependent scour depth around circular bridge piers: Application of ensemble machine learning methods. *Ocean Engineering* **270**, 113611.

Kumar, S., Ojha, C. S. P., Tiwari, N. K. & Ranjan, S. 2023b Exploring the potential of artificial intelligence techniques in prediction of the removal efficiency of vortex tube silt ejector. *International Journal of Sediment Research* **38** (4), 615–627.

Kuncheva, L. I. & Plumpton, C. O. 2010 Choosing parameters for random subspace ensembles for fMRI classification. In: *Multiple Classifier Systems: 9th International Workshop, MCS 2010, Cairo, Egypt, April 7–9, Proceedings 9* (pp. 54–63). Springer, Berlin Heidelberg.

Malik, A., Tikhamarine, Y., Sihag, P., Shahid, S., Jamei, M. & Karbasi, M. 2022 Predicting daily soil temperature at multiple depths using hybrid machine learning models for a semi-arid region in Punjab, India. *Environmental Science and Pollution Research* **29** (47), 71270–71289.

Mamun, A. A., Islam, A. R. M. T., Khosravi, K. & Singh, S. K. 2022 Suspended sediment load prediction using hybrid bagging-based heuristic search algorithm. *Geocarto International* **37** (27), 17068–17095.

Marulasiddappa, S. B., Patil, A. P., Kuntoji, G., Praveen, K. M. & Naganna, S. R. 2023 Prediction of scour depth around bridge abutments using ensemble machine learning models. *Neural Computing and Applications* **36** (1), 1–12.

Minocha, V. K. & Singh, B. 2023 Discussion of 'Prediction of maximum scour depth near spur dikes in uniform bed sediment using stacked generalization ensemble tree-based frameworks' by Manish Pandey, Mehdi Jamei, Masoud Karbasi, Iman Ahmadianfar, and Xuefeng Chu. *Journal of Irrigation and Drainage Engineering* **149** (1), 07022018.

Nasrollahi, A., Ghodsian, M. & Neyshabouri, S. A. A. 2008 Local scour at permeable spur dikes. *Journal of Applied Sciences* **8** (19), 3398–3406.

Nath, D. & Misra, U. K. 2017 Experimental study of local scour around single spur dike in an open channel. *International Research Journal of Engineering and Technology* **4** (6), 2728–2734.

Nguyen, H. & Bui, X. N. 2019 Predicting blast-induced air overpressure: A robust artificial intelligence system based on artificial neural networks and random forest. *Natural Resources Research* **28** (3), 893–907.

Niazkar, M. & Zakwan, M. 2023 Developing ensemble models for estimating sediment loads for different time scales. *Environment, Development and Sustainability* 1–19. doi:10.1007/s10668-023-03263-4.

Nourani, V., Kheiri, A. & Behfar, N. 2022 Multi-station artificial intelligence based ensemble modeling of suspended sediment load. *Water Supply* **22** (1), 707–733.

Özyaman, C., Yerdelen, C., Eris, E. & Daneshfaraz, R. 2022 Experimental investigation of scouring around a single spur under clear water conditions. *Water Supply* **22** (3), 3484–3497.

Pande, C. B., Costache, R., Sammen, S. S., Noor, R. & Elbeltagi, A. 2023 Combination of data-driven models and best subset regression for predicting the standardized precipitation index (SPI) at the Upper Godavari Basin in India. *Theoretical and Applied Climatology* **152** (1–2), 535–558.

Pandey, M., Ahmad, Z. & Sharma, P. K. 2016 Estimation of maximum scour depth near a spur dike. *Canadian Journal of Civil Engineering* **43** (3), 270–278.

Pandey, M., Ahmad, Z. & Sharma, P. K. 2018 Scour around impermeable spur dikes: A review. *ISH Journal of Hydraulic Engineering* **24** (1), 25–44.

Pandey, M., Jamei, M., Karbasi, M., Ahmadianfar, I. & Chu, X. 2021 Prediction of maximum scour depth near spur dikes in uniform bed sediment using stacked generalization ensemble tree-based frameworks. *Journal of Irrigation and Drainage Engineering* **147** (11), 04021050.

Pandey, M., Karbasi, M., Jamei, M., Malik, A. & Pu, J. H. 2023 A comprehensive experimental and computational investigation on estimation of scour depth at bridge abutment: Emerging ensemble intelligent systems. *Water Resources Management* **37**, 3745–3767.

Patel, H. K., Arora, S., Lade, A. D., Kumar, B. & Azamathulla, H. M. 2023 Flow behavior concerning bank stability in the presence of spur dike – a review. *Water Supply* **23** (1), 237–258.

Pham, Q. B., Kumar, M., Di Nunno, F., Elbeltagi, A., Granata, F., Islam, A. R. M. T., Talukdar, S., Nguyen, X. C., Ahmed, A. N. & Anh, D. T. 2022 Groundwater level prediction using machine learning algorithms in a drought-prone area. *Neural Computing and Applications* **34** (13), 10751–10773.

Pourshahbaz, H., Abbasi, S., Pandey, M., Pu, J. H., Taghvaei, P. & Tofangdar, N. 2022 Morphology and hydrodynamics numerical simulation around groynes. *ISH Journal of Hydraulic Engineering* **28** (1), 53–61.

Richardson, E. V. & Davis, S. R. 2001 *Evaluating Scour at Bridge*. Federal Highway Administration, Office of Bridge Technology, McLean, VA.

Rodriguez-Lozano, F. J., León-García, F., Gámez-Granados, J. C., Palomares, J. M. & Olivares, J. 2020 Benefits of ensemble models in road pavement cracking classification. *Computer-Aided Civil and Infrastructure Engineering* **35** (11), 1194–1208.

Salmasi, F., Sihag, P., Abraham, J. & Nouri, M. 2023 Experimental investigation and prediction of free fall jet scouring using machine learning models. *International Journal of Sediment Research* **38** (3), 405–420.

Sepahvand, A., Singh, B., Sihag, P., Nazari Samani, A., Ahmadi, H. & Fiz Nia, S. 2021 Assessment of the various soft computing techniques to predict sodium absorption ratio (SAR). *ISH Journal of Hydraulic Engineering* **27** (sup1), 124–135.

Sharafati, A., Haji Seyed Asadollah, S. B., Motta, D. & Yaseen, Z. M. 2020 Application of newly developed ensemble machine learning models for daily suspended sediment load prediction and related uncertainty analysis. *Hydrological Sciences Journal* **65** (12), 2022–2042.

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2024.025/1366312/wst2024025.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

corrected Proof

Singh, B. 2020 Prediction of the sodium absorption ratio using data-driven models: A case study in Iran. *Geology, Ecology, and Landscapes* **4** (1), 1–10.

Singh, N. B., Devi, T. T. & Kumar, B. 2022 The local scour around bridge piers – a review of remedial techniques. *ISH Journal of Hydraulic Engineering* **28** (sup1), 527–540.

Skurichina, M. & Duin, R. P. 2002 Bagging, boosting, and the random subspace method for linear classifiers. *Pattern Analysis & Applications* **5**, 121–135.

Tao, D., Tang, X., Li, X. & Wu, X. 2006 Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28** (7), 1088–1099.

Taylor, K. E. 2001 Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* **106** (D7), 7183–7192.

Tripathi, R. P. & Pandey, K. 2023 Gene-expression programming for scour around spur dike. *International Journal of Hydrology Science and Technology* **15** (3), 295–303.

van der Wal, M. 2020 Bank protection structures along the Brahmaputra-Jamuna River, a study of flow slides. *Water* **12** (9), 2588.

Williamson, D. F., Parker, R. A. & Kendrick, J. S. 1989 The box plot: A simple visual method to interpret data. *Annals of Internal Medicine* **110** (11), 916–921.

Yeganeh-Bakhtiary, A., Houshangi, H., Hajivalie, F. & Abolfathi, S. 2017 A numerical study on hydrodynamics of standing waves in front of Caisson breakwaters with WCSPH model. *Coastal Engineering Journal* **59** (1), 1750005–1.

Yeganeh-Bakhtiary, A., Houshangi, H. & Abolfathi, S. 2020 Lagrangian two-phase flow modeling of scour in front of vertical breakwater. *Coastal Engineering Journal* **62** (2), 252–266.

Yeganeh-Bakhtiary, A., EyvazOghli, H., Shabakhty, N., Kamranzad, B. & Abolfathi, S. 2022 Machine learning as a downscaling approach for prediction of wind characteristics under future climate change scenarios. *Complexity* **2022**. doi:10.1155/2022/8451812

Yeganeh-Bakhtiary, A., EyvazOghli, H., Shabakhty, N. & Abolfathi, S. 2023 Machine learning prediction of wave characteristics: Comparison between semi-empirical approaches and DT model. *Ocean Engineering* **286**, 115583.

Zaghloul, N. A. 1983 Local scour around spur-dikes. *Journal of Hydrology* **60** (1–4), 123–140.

Zhang, H. & Nakagawa, H. 2009 Characteristics of local flow and bed deformation at impermeable and permeable spur dykes. *Annual Journal of Hydraulic Engineering, JSCE* **53**, 145–150.

First received 2 October 2023; accepted in revised form 4 January 2024. Available online 29 January 2024

Downloaded from http://iwaponline.com/wst/article-pdf/doi/10.2166/wst.2024.025/1366312/wst2024025.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Automatic Construction of Interval-Valued Fuzzy Hindi WordNet using Lexico-Syntactic Patterns and Word Embeddings

MINNI JAIN

Computer Science & Engineering Department, Delhi Technological University (DTU), Delhi, India (e-mail: minnijain@dtu.ac.in).

RAJNI JINDAL

Computer Science & Engineering Department, Delhi Technological University (DTU), Delhi, India (e-mail: rajnijindal@dtu.ac.in).

AMITA JAIN

Netaji Subhas University of Technology (NSUT), East Campus, Delhi, India (e-mail: amita.jain@nsut.ac.in).

A computational lexicon is the backbone of any language processing system. It helps computers to understand the language complexity as a human does by inculcating words and their semantic associations. Manually constructed famous Hindi WordNet (HWN) consists of various classical semantic relations (crisp relations). To handle uncertainty and represent Hindi WordNet more semantically, Type- 1 fuzzy graphs are applied to relations of Hindi WordNet. But uncertainty in the crisp membership degree is not considered in Type 1 fuzzy set (T1FS). Also collecting billions (5,55,69,51,753 relations in HWN) of membership values from experts (humans) is not feasible. This paper applied the concept of Interval-Valued Fuzzy graphs and proposed Interval- Valued Fuzzy Hindi WordNet (IVFHWN). IVFHWN automatically identifies Interval- Valued Fuzzy relations between words and their degree of membership using word embeddings and lexico-syntactic patterns. The experimental results for the word sense disambiguation problem show better outcomes when IVFHWN is being used in place of Type 1 Fuzzy Hindi WordNet and classical Hindi WordNet.

Index Terms Hindi WordNet, Interval- Valued Fuzzy Graphs, Computational Lexicon, Natural Language Processing, Low-Resource Language

1 INTRODUCTION

Human mind is trained from childhood to handle the language complexities. But for computers to comprehend multiple meanings and context precisely is an arduous process. To make the language understandable by computers, it is mandatory to comprehend the words and their association with other words. To represent this semantic association, researchers created lexicons. WordNet [1] is a lexicon that is available for almost all languages and extensively being used for many natural language processing applications. WordNet for Hindi, called *Hindi WordNet*, having 16 relations between words, is developed by the Indian Institute of Bombay (IIT-B) in early 2000 [2].

The Hindi WordNet (HWN) is an available online system that comprises various semantic and lexical relations between the Hindi words. It has been explored for almost every Hindi language processing application, such as Word processing Information Retrieval (IR), Question Answering systems (QRs), Machine Translation (MT), etc. [3-6]. HWN also has been used as a base resource to develop new lexicons such as Hindi SentiWordNet [7]. Hindi WordNet has 1.05,423 unique words and 40465 synsets with various crisp relations between the words[1]. However, in real-life situations, relationships between words (as used to convey natural language) are not crisp; rather, they are a matter of degree, with a gradual transition from being related to not being related. For instance, as shown in Fig. 1 (a), there is a “meronymy” semantic relation between फल and गुठली. In classical Hindi WordNet, all relations are crisp, but it is easy to observe that while some fruits (such as bananas) don't have गुठली, the majority of them have. Words in natural language have partial relations, which crisp classical Hindi WordNet does not handle. Fig. 1(b) shows fuzzified relations of Hindi WordNet.

In this direction, Jain & Lobiyal investigated this concern and proposed Fuzzy Hindi WordNet (Type-1 Fuzzy) that comprises fuzzified relations [4]. This was the first work to incorporate real-life uncertainties in Hindi WordNet. The authors proved the better performance of fuzzy Hindi WordNet over classical Hindi WordNet by handling word sense disambiguation. Jain et al. proposed a very compelling idea of Fuzzy Hindi WordNet (FHWN), and they assigned the link strengths between two words in the FHWN network by linguistic experts. However, the concept of involving experts is impractical when we want to incorporate a large number of the relation between words. For instance, Hindi WordNet has 1,05,423 unique

[1] https://www.cfilt.iitb.ac.in/wordnet/webhwn/wn.php

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM 2375-4699/2024/02-ART
http://dx.doi.org/10.1145/3643132

Just Accepted

words/nodes, and the possible relation between these unique words will be ${}^{n}C_2$ = 5556951753 relations, where n is the unique words. It is next to impossible to assign membership value to this large number of relations with the help of experts. Their idea is very significant to express the relations between words according to real-world scenarios, but it is virtual until researchers can use it for NLP applications. In this direction, it is desirable to generate fuzzy Hindi WordNet relations automatically. One more issue in the literature is that the authors considered average of the fuzzy membership values taken from experts and converted into a single value, which results in loss of opinion and leads to a loss in the valuable data. This is again not handled by existing work.

The proposed work overcomes the shortcomings as mentioned above in the following ways:

1. Overcome the requirement of enormous human resources (linguistic experts) for membership value assignment: The work proposed Interval- Valued Fuzzy Hindi WordNet (IVFHWN) that automatically identifies Interval- Valued Fuzzy semantic relations between Hindi words using lexico-syntactic patterns and various Word embeddings. We render IVFHW as Interval- Valued graph where nodes/vertices represent the Hindi words and edges represent the IVF relations between words. Fig. 1 (b) shows the IVFHW relations for the word 'फल'. The words "फल and पेड" are connected to each other with IVF holonymy relation having [0.66, 0.75] membership value. Lexico-syntactic patterns are generalised linguistics structures that used to extract semantic relationships between words.
2. Handles the uncertainty in the Type-1 fuzzy membership degree. In this paper, we have incorporated Interval-Valued Fuzzy relations, a generalised form of Type-2 Fuzzy relations. The motivation behind using IVFs is uncertainty in the membership degree of Type-1Fuzzy. This part also resolved the issue of loss of information occurs in averaging values taken by different experts.

To evaluate the proposed Interval- Valued Fuzzy Hindi WordNet, we used a heath domain corpus publicly available at the IIT Bombay website [8]. In order to assess the proposed method, the mean square error (MSE) is calculated between the expert value and the proposed value. To arrive at the expected value, we enlisted the help of a group of famous lexicographers (linguists who specialise in Hindi) and fuzzy logic experts. The results demonstrated proposed method (IVFHWN) is providing good results. The proposed method can make the concept of fuzzifying Hindi WordNet feasible for all applications of natural language processing. The proposed system is also compared with Hindi WordNet, Fuzzy Hindi WordNet (where membership values are taken from linguistic experts). All three lexicons (HWN, FHWN and proposed IVFHWN) are applied to the health sense dataset for word sense disambiguation. The results clearly show that IVFHW outperforms the remaining two.

The main contributions of the proposed work are as follows:

1. The proposed method constructed an Interval-Valued Fuzzy Hindi WordNet that handles the uncertainties of classical Hindi WordNet (Crisp) and Type-1 Fuzzy Hindi WordNet.
2. In contrast to state-of-the-art methods, which rely on manual generation, this work represents the first automated approach to identifying relationships and determining IVF membership values between words.
3. For the first time, lexico-syntactic patterns and word embeddings are used to learn semantic relatedness between words which helped to represent the computational lexicon more semantically.
4. The interval-valued fuzzy graph is used for the first time to represent the computational lexicon for Hindi language.

After the introduction in Section 1, a summary of the background is offered in Section 2. Preliminaries, which provide fundamental concepts, terminologies, and mathematical terms necessary to comprehend the work in Section 3. Section 4 deals with the proposed work, which consists of the definitions of Interval-Valued Fuzzy Hindi WordNet and relations with their identification and membership degree assignment methods. Section 5 describes the experimental setup, discussion of the results and comparison of the available systems. Finally, in Section 6, the conclusion and ideas for future work are presented.

Just Accepted

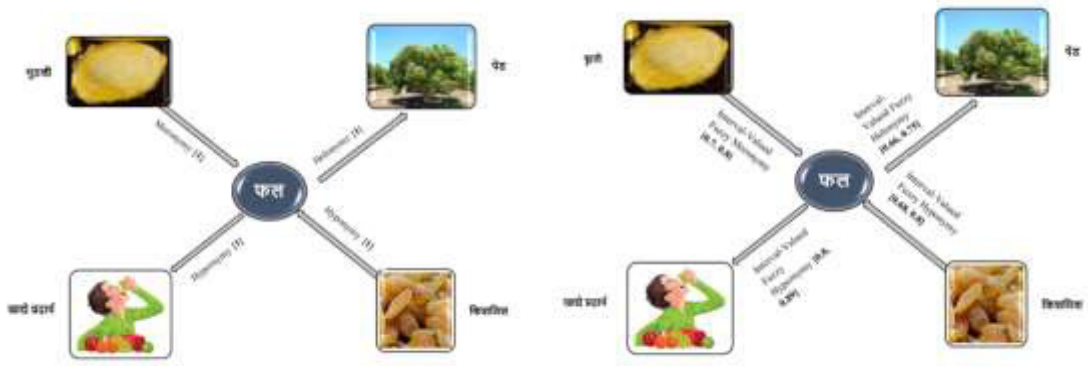


Fig. 1(a) Crisp relations for the word 'फल' in Hindi WordNet Fig. 1(b) Interval-valued Fuzzy relations for the word 'फल' in proposed IVFHW

2 BACKGROUND

In this section, we provide a brief background of all the concepts used in our proposed work.

With fuzzy sets, human notions like "big", "small", "tall', "wide", etc. can be translated into a form usable by computers. The fuzzy set theory [9] has been widely applied to incorporate the imprecision and uncertainty in various areas, including natural language [10-15]. Lexicons play an essential role in natural language understanding. So applying fuzzy set theory on widely used lexicon can improve the performance of lexicon-based NLP applications. In the early nineties, the author offered space as a network of fuzzy relationships called fuzzy concept networks by introducing four types of fuzzy relationships between the concept that are a positive association, negative association, specialisation and generalisation [15].

Further, this study of Concept networks was used by Shyi Ming Chen et al. [11] to enhance information retrieval performance. Since constructing the Fuzzy concept network manually was a tedious task, a method to automatically construct the fuzzy concept network was developed [12]. In 2015 Jain et al. got inspired by the idea of fuzzy concept network [4]. They proposed a fuzzy Hindi WordNet where authors assigned fuzzy membership value to the seventeen relations available in crisp Hindi WordNet. They assigned membership value with the help of various experts. The authors also show the performance of proposed Hindi fuzzy centrality measures and Hindi WordNet by applying on word sense disambiguation using various graph centrality measures. In Jain et al., authors applied the concept of Fuzzy Hindi WordNet on Hindi query expansion which also disambiguates the ambiguous query. Results of their work show that Fuzzy WordNet outperforms the crisp WordNet in the field of query expansion, too [16].

WordNet is available for several languages, including English, Hindi, Swedish, Portuguese etc. Many authors have applied the concept of fuzzifying the components of WordNet (nodes, edges, synsets). For instance, in 2009, authors presented an algorithm to construct Swedish fuzzy synsets [48] and in 2011. The authors proposed Portuguese fuzzy synsets [47]. Similarly, in 2021, Yousef et al. suggested interval probabilistic fuzzy WordNet for the English language [50]. They again fuzzified synsets of WordNet-based on the probabilistic fuzzy concept. Work by Sonakshi et al. in 2018 provides fixed weight to relations in WordNet for the English language [49]. Assigning a crisp membership degree does not account for uncertainty associated with methods by which membership values are computed. However, the proposed work is the first work that uses Interval-Valued Fuzzy (generalise form of type -2 fuzzy logic) to create IVF WordNet for the Hindi language. IVFHWN automatically assigns relations and interval membership between words using lexico-syntactic patterns and Hindi word embeddings.

Word embeddings were first introduced in 2003 [17], and after that, several developments have occurred in this field. Word2Vec [18] presented the method for word vectors training. The embeddings models familiarised by them established a new state-of-art for many tasks such as Name Entity Recognition, Word Sense Disambiguation etc. Mikolov et al. [18] results are further improved by GloVE [19] and FastText [20]. GloVE utilised a co-occurrence matrix, whereas FastText used the sub-word information to create word vectors. Later, BERT [21] applied the transformer architecture to establish new a state-of-the-art across various tasks. It was able to learn deep bidirectional context instead of just two unidirectional contexts, which helped it outperform previous models. The authorised repository for FastText has numerous pre-trained word embedding for multiple languages, including French, Hindi, Polish etc.

Just Accepted

Many researchers exploited word embeddings for Hindi language processing, such as employing Word Embeddings and WordNet-based Features to recognise Multiword Expressions in Hindi. Pujari et al. used word embeddings for debiasing gender-biased Hindi words. The authors used word2vec specifically to propose a debiasing algorithm [22]. By Sirigiri et al. word2vec's continuous bag-of-word model (CBOW) was utilised to present a system for resolving spelling errors in Hindi text obtained by OCR. [23]. Recently, Hindi BERT and Hindi Fasttext models are exploded to detect Covid-19 related fake news and abusive content on social media [24]. Word2Vec and FastText word embeddings are also used in the Hindi Sentiment Analysis area for feature generation [25]. In 2019, authors proposed a deep learning approach for Name Entity Recognition (NER) for Hindi Text, and they used Word2Vec, FastText and Bert for the feature extraction [26]. Kumari et al. investigated various word embedding approaches for word sense disambiguation in the Hindi language. They tested the quality of the obtained word embeddings using Pearson correlation and utilised an existing word2vec model to create Hindi word embeddings using Wikipedia articles [27]. From the literature review, it is visible that Word embeddings have been exploited for almost every problem of Hindi language processing. [28 [29]

Hindi WordNet was built by researchers at IIT Bombay and is now widely utilised in numerous NLP applications. In the Hindi language, the Hindi WordNet is employed to resolve ambiguity [30-32]. Hindi WordNet is also applied for query expansion and optimisation in Hindi [33, 39]. The researchers have optimised information retrieval by using Ontology in Hindi WordNet [34]. The authors used Wikipedia to construct synsets in WordNets for Indian languages automatically [35]. Authors recently used Hindi WordNet to retrieve soft clusters of web search results for a Hindi query [36]. For the detection and repair of spelling errors, Jain et al. employed Hindi WordNet [3].

After going through the literature, we find that the Hindi WordNet is an essential tool, and it has been exploited and by many researchers. Researchers also keep improving WordNet by fuzzify its different components like nodes, relation and synsets, But researchers have not explored the Interval-valued fuzzy for Hindi WordNet that uses word embeddings for automatically assigning membership values. It reduces the human resources and uncertainty of crisp membership degree.

3 PRELIMINARIES

This section briefly introduces some key concepts and terminologies with the mathematical terms required to understand the approach.

3.1 Interval-Valued Fuzzy Set

An Interval-Valued Fuzzy set, for an element x in the universe of discourse *X*, can be defined as [37],

$$\tilde{A} = \left\{\left(x, \left[\mu_{\tilde{A}_L}, \mu_{\tilde{A}_U}\right]\right) | x \in X\right\} \quad (1)$$

Where, $\left[\mu_{\tilde{A}_L}, \mu_{\tilde{A}_U}\right]$ is an interval between [0,1]. For all *x*, are the membership interval iff:

$$0 \leq \mu_{\tilde{A}_L} \leq \mu_{\tilde{A}_U} \leq 1$$

Fig. 2. Illustrate the membership value at *x'* of the Interval-Valued Fuzzy Set Ã; thereby, the membership value of *x'* is the interval of $[(\mu_{\tilde{A}_L}, \mu_{\tilde{A}_U})]$

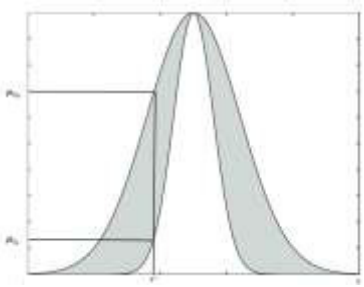


Fig.2. Interval-Valued Fuzzy Set Ã

Just Accepted

3.2 Interval-Valued Fuzzy Graph

An Interval- Valued fuzzy graph $G^* = (V, E)$ is a pair $G^* = (\tilde{A}, \tilde{B})$ where $\tilde{A} = [\mu_{\tilde{A}_L}, \mu_{\tilde{A}_U}]$ is an Interval- Valued Fuzzy set on V with $0 \leq \mu_{\tilde{A}_L}, \leq \mu_{\tilde{A}_U} \leq 1$and $\tilde{B} = [\mu_{\tilde{B}_L}, \mu_{\tilde{B}_U}]$ is an Interval- Valued fuzzy relation E with $0 \leq \mu_{\tilde{B}_L}, \leq \mu_{\tilde{B}_U} \leq 1$. [38,43]

Example: Consider a graph $G^* = (V, E)$ such that $V = \{x,y,z\}$ and $E = \{xy, yz, zx\}$. Let $\tilde{A} = [\mu_{\tilde{A}_L}, \mu_{\tilde{A}_U}]$ be an Interval- Valued Fuzzy set on V *(vertex)* and Let $\tilde{B} = [\mu_{\tilde{B}_L}, \mu_{\tilde{B}_U}]$ be an Interval- Valued Fuzzy relation E *(edges)*. Matrices for V and E are:

$[\mu_{\tilde{A}_L}, \mu_{\tilde{A}_U}]$	*x*	*y*	*z*
	[0.3,0.4]	[0.5, 0.7]	[0.4,0.6]

$[\mu_{\tilde{B}_L}, \mu_{\tilde{B}_U}]$	*xy*	*yz*	*zx*
	[0.8,0.9]	[0.4, 0.7]	[0.3,0.4]

3.3 Hindi WordNet (HWN)

Among the various Indian language WordNets, the Hindi WN is the first WordNet that came into existence from 2000 onwards. HWN is inspired by English WN, and it is created manually using lexical knowledge acquired from various dictionaries. It comprises nouns, adjectives, verbs and adverbs structured into synonym sets, each demonstrating one lexical concept. HWN is used in various natural language processing applications like Word Sense Disambiguation [4], Sentiment Analysis and Opinion Mining [51], Information Retrieval [16] and many more. WordNet is a word sense network that can be represented as a graph also. A word sense node in this network is a synset. Each synset is linked with other synsets through lexical and semantic relations like hyponymy, hypernymy, troponymy, entailment, antonymy etc. Hindi WordNet consists of 17 relations, including same part-of-speech and cross part-of-speech relations.

3.4 Word Embeddings or Word Vectors

Word Embedding or Word Vector is a recent development in computational linguistics. Many deep learning-based natural language processing applications use it as a fundamental resource. A word embedding is a vector representation of a word with the property that the embeddings of comparable words in an N-dimensional vector space are close to each other. Word2Vec is a common technique for learning word embeddings that employ shallow neural networks. [11]. Cosine similarity is a metric to measure the distance between two words. Word2Vec can be used to find the belongingness between two words in a fuzzy graph.

3.5 ConceptNet

ConceptNet [42] is a knowledge base that can be visualised as a graph that connects natural language words or phrases (concepts) with tagged edges. The relationships between nodes in conceptNet are gathered from a variety of sources, including crowdsourcing, Open Mind Common Sense (OMCS) data, Games with a Purpose, and expert-created resources. ConceptNet is aimed to reflect the commonsense to comprehend better the meaning behind the words people use. It comprises about 21 million edges and over 8 million nodes. It supports 83 languages, including Hindi. That's why in this work, conceptNet can be used easily because it supports the Hindi language. But vocabulary size for the Hindi language is 21163 only, which is not sufficient to use it alone for relation identification and assignment of membership value. Currently there are 34 relations exists in conceptNet including MotivatedByGoal, UsedFor, SymbolOf, Antonym, SenseOf, DistinctFrom, ReceivesAction, EtymologicallyRelatedTo, PartOf, LocatedNear, ObstryctedBy, RelatedTo, MotivatedByGoal, SimilarTo, MannerOf, Synonymy, MadeOf, AtLocation, HasProperty etc..

4 PROPOSED INTERVAL-VALUED FUZZY HINDI WORDNET (IVFHW)

In the following section, we define Interval-Valued Fuzzy Hindi WordNet (IVFHWN) and various IVF relationships. This section also comprises the identification of IVF relationships and automatic assignment of membership value to the IVF relationships. These relationships are identified using various lexico-syntactic patterns and If-Then rules, whereas different word embeddings are used to assign membership values automatically.

Just Accepted

4.1 Definition of Interval-Valued Fuzzy Hindi WordNet and Its relations

Definition A. An Interval-Valued Fuzzy Hindi WordNet Graph is denoted as IVFHWN (σ, $\mu_{\tilde{r}}$), which is a pair of functions σ : W→{1} and $\mu_{\tilde{r}}$: W×W→ $[0 \leq \mu_{\tilde{r}_L} \leq \mu_{\tilde{r}_U} \leq 1]$, where σ is a set of nodes representing the set of Hindi open class words (set of synsets) with the unity strength of every node. It means every Hindi word entirely belongs to Interval- Valued Fuzzy Hindi WordNet. $\mu_{\tilde{r}}$, which is Interval- Valued Fuzzy membership degree denotes the strength of an edge between two Hindi words, where $\mu_{\tilde{r}_L} and\ \mu_{\tilde{r}_U}$ Indicates the lower and upper membership value of relation r̃, respectively.

For example, Fig. 3 shows an Interval- Valued Fuzzy Hindi WordNet graph where w_1, w_2, w_3,..... w_5 are the Hindi words; word w_1 is linked with word w_2 with hypernymy ($\widetilde{hr}$) relationship having IV fuzzy degree of strength $\mu_{\widetilde{hr}} \in$ [0.2,0.4]; similarly, word w_2 is connected with word w_3 with Meronymy ($\tilde{M}$) relationship having IV fuzzy degree of strength $\mu_{\tilde{M}} \in$ [0.3,0.4] and so on.

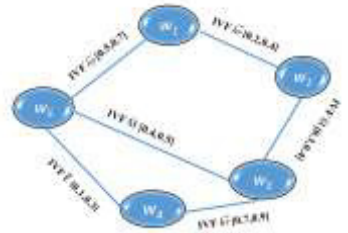


Fig. 3.: Interval-Valued Fuzzy Graph

In the Interval-Valued Fuzzy WordNet network, there are sixteen fuzzy relationships (same part-of-speech and cross part-of-speech) between words, and they are defined formally as follows:

Definition B: Interval-Valued Fuzzy Association (Ãs) relationship represented as (w_1, w_2, $\mu_{\tilde{A}s}$) and defined as IVF relation between w_1 and w_2 (two synsets) where w_1, $w_2 \in W$ and $\mu_{\tilde{A}s}$: $W\times\ W \rightarrow [\mu_{\tilde{A}s_L}(x), \mu_{\tilde{A}s_U}(x)]$, such that w_1 and w_2 are approximately synonymy where W denotes to the set of all synsets, $\mu_{\tilde{A}s}$ indicating the degree of strength of the relationship with $[\mu_{\tilde{A}s_L}(x),\ \mu_{\tilde{A}s_U}(x) \in [0,1]$ indicating the lower and upper interval of $\mu_{\tilde{A}s}$. IVF association (Ãs) represents the partial similar meaning between two words, for instance (सोफा, कुर्सी, [0.6,0.75]) where interval [0.6,0.75] shows the strength of the relationship in the form of interval. More examples are: (मतदान, राय, [0.4,0.6]) (दस्ता, कॉपी, [0.54, 0.72]) (मेज, स्टूल, [0.5,0.7])

Definition C: Interval-Valued Fuzzy Hypernymy ($\widetilde{hr}$) and Hyponymy ($\widetilde{hp}$) can be represented as (w_1, w_2, $\mu_{\widetilde{hr}}$ / $\mu_{\widetilde{hp}}$) and defined as IVF relation between w_1 and w_2 (two synsets) where w_1, $w_2 \in W$ and $\mu_{\widetilde{hr}}$: $W\times\ W \rightarrow [\mu_{\widetilde{hr}_L}(x),\ \mu_{\widetilde{hr}_U}(x)]$, $\mu_{\widetilde{hp}}$: $W\times\ W \rightarrow [\mu_{\widetilde{hp}_L}(x),\ \mu_{\widetilde{hp}_U}(x)]$, such that w_1 is an approximate subset/superset of w_2 where W denotes the set of all synsets, $\mu_{\widetilde{hr}}$ / $\mu_{\widetilde{hp}}$ indicating the degree of strength of the relationship with $\mu_{\widetilde{hr}_L}(x)$, $\mu_{\widetilde{hr}_U}(x)$, $\mu_{\widetilde{hp}_L}(x)$, $\mu_{\widetilde{hp}_U}(x)$ indicating the lower and upper interval value of $\mu_{\widetilde{hr}}$/ $\mu_{\widetilde{hp}}$.

For instance, there is IVF hypernymy/hyponymy relation between (नेल कटर and हथियार with interval membership value [0.6, 0.8] where 0.6 is the lower interval value and 0.8 is the upper interval value. Similarly, (बईमान and दुर्जन are related with strength of [0.4, 0.5]).

Definition D: Interval-Valued Fuzzy Meronymy ($\tilde{M}$) and Holonymy ($\tilde{H}o$) are semantic relations between two synsets w_1 and w_2 and used to capture the part-whole relationships between w_1, w_2, where w_1, $w_2 \in W$ and $\mu_{\tilde{M}}$: $W\times\ W \rightarrow [\mu_{\tilde{M}_L}(x),\ \mu_{\tilde{M}_U}(x)]$, $\mu_{\tilde{H}o}$: $W\times\ W \rightarrow [\mu_{\tilde{H}o_L}(x),\ \mu_{\tilde{H}o_U}(x)]$, where W denotes the set of all synsets, $\mu_{\tilde{M}}$ / $\mu_{\tilde{H}o}$ indicating the degree of strength of the relationship. For example, (बटन, कुर्ता [0.8,0.9]), (फूल, पेड़ [0.65, 0.74]), (बिजली बारिश [0.5, 0.66])

Definition E. Interval-Valued Fuzzy Antonyms (Ã): It is a lexical relation between two words with approximately opposite meanings in some context. Interval- Valued Fuzzy Antonyms can be represented as (w_1, w_2, $\mu_{\tilde{A}}$) where $\mu_{\tilde{A}}$: $W\times\ W \rightarrow [\mu_{\tilde{A}_L}(x),\ \mu_{\tilde{A}_U}(x)]$ and w_1, $w_2 \in W$. Illustration: (बेटी, लड़का [0.6, 0.78]), (शाम, दिन [0.7, 0.8]) and (मुर्दा बेहोश [0.6, 0.73])

Definition F. Interval-Valued Fuzzy Entailment (Ẽ): It is a one-way semantic relation that holds between two synsets of verbs if the truth of verb y partially follows the truth of verb x. It can be repressed as $v_1 \rightarrow v_2$ having IVF membership $\mu_{\tilde{E}}$: $V\times\ V \rightarrow [\mu_{\tilde{E}_L}(x),\ \mu_{\tilde{E}_U}(x)]$, v1, v2 ∈ V and V (V⊂W) is the set of all verbs in the Hindi language.

Just Accepted

For instance, (सब्जी काटना, खाना बनाना [0.5, 0.62]) , (निरोगी होना, व्यायाम करना [0.8, 0.9]), (बुरी आदते होना, बुरे संस्कार होना [0.4,0.54])

Definition G. Interval-Valued Fuzzy Troponymy ($\tilde{T}$): This relation relates two synsets of verbs, verb v_A denoting elaboration of verb v_B in an approximately specific manner. Interval- Valued Fuzzy Troponymy can be represented as (v_A, v_B, $\mu_{\tilde{T}}$), where v_A, $v_B \in V$and $V \subset W$is the set of verbs in Hindi Language. and $\mu_{\tilde{T}}$: $V \times V \rightarrow [\mu_{\tilde{T}_L}(x), \mu_{\tilde{T}_U}(x)]$ is the Interval- Valued membership value for fuzzy tropynymy. Illustration: (भागना, चलना [0.6, 0.69]), (हसना, खुश होना [0.8, 0.9]) and (सिलाई करना, सीखना [0.5,0.63])

Definition H. Interval-Valued Fuzzy Causative ($\tilde{C}$) is a lexical relation which associates the causative verbs and shows interdependencies between them. For this relation, the membership value is always unity because it shows the relation between the different morphological forms of a verb. For Illustration **(**पीना पिलाना, 1) and (सोना सुलाना , 1)

Definition I: Interval-Valued Fuzzy Ability Link ($\tilde{A}l$) is semantic relation, represented as (n. v, $\mu_{\tilde{A}l}$), where $n \in N$, $v \in V$and N, $V \subset W$is the set of nouns and verbs in Hindi Language respectively. $\mu_{\tilde{A}l}$ is the Interval- Valued membership value for the fuzzy ability link. This relation specifies the inherited feature of a nominal word. For Instance (पंछी, उड़ना [0.7, 0.8])

Definition J: Interval- Valued Fuzzy Capability Link ($\tilde{C}_L$) is semantic relation, represented as (n. v, $\mu_{\tilde{C}_L}$), where $n \in N$, $v \in V$and N, $V \subset W$is the set of nouns and verbs in Hindi Language respectively, and $\mu_{\tilde{C}_L}$ is the Interval-Valued membership value for fuzzy capability link. This relation specifies the developed feature of a nominal word. For instance, the capability of नौकरी करना of a आदमी is more than the capability of नौकरी करना of a लड़का because in most of the cases आदमी refers to an adult whereas लड़का refers to a juvenile. This relation can be represented as: (आदमी, नौकरी करना [0.9, 0.98]) (लड़का, नौकरी करना[0.5,0.6])

Definition K: Interval- Valued Fuzzy Function Link ($\tilde{F}_L$) is represented as (n. v, $\mu_{\tilde{F}_L}$), where $n \in N$, $v \in V$and N, $V \subset W$is the set of nouns and verbs in Hindi Language respectively, and $\mu_{\tilde{F}_L}$ is the Interval- Valued membership value for fuzzy function link. It is a semantic relation that specifies the function of a nominal word. This relation can be understood by an example like: "पढ़ाना is a function of a अध्यापक ". In an ideal world, all teachers would be expected to teach; nevertheless, many administrative professors (such as the Vice-Chancellor of a university) are not allocated any teaching responsibilities. In addition, some teachers are given a heavier teaching load than others. A professor, for example, is given a lower teaching load than an assistant professor or associate professor. Professors, on the other hand, are given more administrative and research responsibilities (as well as a lower teaching load) than assistant and associate professors. The strength of the respective function is used to determine the weight of a fuzzy function connection.

Definition L: Interval- Valued Fuzzy Attribute ($\tilde{Att}$) is a semantic relationship between the synset of a noun and adjective. It represents the partial properties of a noun. It can be represented as (n. a_d, $\mu_{\tilde{Att}}$), where $n \in N$, $a_d \in A_d$ and N, $A_d \subset W$is the set of nouns and adjectives in Hindi Language, respectively, and $\mu_{\tilde{Att}}$ is the Interval- Valued membership value with upper and lower intervals. For instances (सुशील ,लड़की [0.5, 0.6]), (बंजर, भूमि [0.7,0.78])

Definition M: Interval- Valued Fuzzy Modifies Noun ($\tilde{M}_N$) is the semantic relationship between adjectives and nouns. It comprises cases when adjectives can only go with specific nouns. For this relation, the membership value is always unity. It can be represented as (n. a_d, $\mu_{\tilde{M}_N}$), where $n \in N$, $a_d \in A_d$ and N, $A_d \subset W$ is the set of nouns and adjectives in Hindi Language respectively, and $\mu_{\tilde{M}_N} = \{1\}$ is the Interval- Valued membership value. For example (संस्कारी, इंसान , 1)

Definition N: Interval- Valued Fuzzy Modifies Verb ($\tilde{M}_v$) is the semantic relation between verb and adverb such that an adverb modifies the specific verb. For this relation, the membership value is always unity. It means this is always a classic relation. This relation is anti-symmetric, anti-reflexive, and max-*-nontransitive. For instance (तेज ,चलना, 1)

Definition O: Interval- Valued Fuzzy Derived From ($\tilde{D}_F$) specifies the origin (root) form from which a word is derived. It is a lexical relation, and the membership value for this relation is always unity. This relation is anti-symmetric, anti-reflexive, and max-*-nontransitive. For instance (सुंदरता, सुंदरी, सुंदर , 1)

Just Accepted

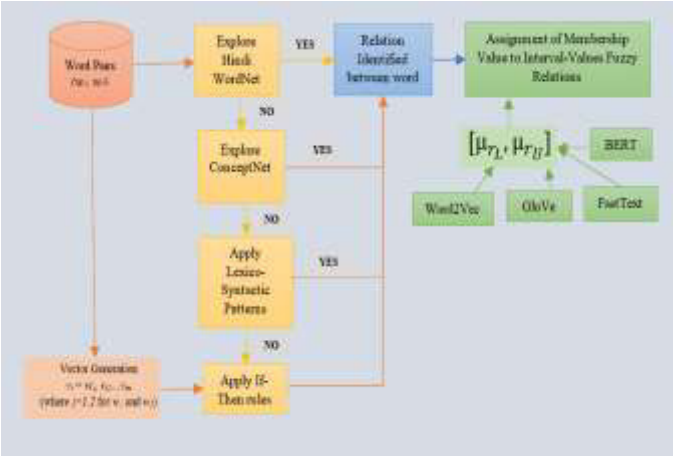


Fig. 4. Work Flow Diagram of the Proposed IVFHW

4.2 Identification of Interval-Valued Fuzzy relations

This section discusses the lexico-syntactic patterns and If-then rules for the identification of IVF relationships between words. All the word pairs are first explored in Hindi WordNet; if the relation is found in WordNet, it goes to the next step: the assignment of membership value, or else word pairs are searched in conceptNet. It not available in conceptNet also, then below discussed lexico-syntactic patterns and if-then rules are used to find the relationship. The overall architecture is shown in fig. 4.

4.2.1 Identification of IVF association relation includes the following steps:

a) Let say w1 and w2 represent words, and the objective is to identify IVF Ãs the relation between them, for this, the first step is to generate vector vi = vi1, vi2...vim (where i=1,2 for w1 and w2). Vector v1 and vector v2 comprise co-occurred VP and NP and AdjP. For instance, vectors for words (स्कू टर and मोटरसाइकल) are shown in fig.5. Vectors are generated by considering all the terms that co-occur with a specific word in a corpus. The next step is calculating the weight of each term in a vector using normalize Tf×idf (Term frequency multiplied by inverse document frequency) method shown in formula (2).

$$\mathrm{Tf} \times \mathrm{idf}\,(t, d_i) = \frac{(0.5+0.5\,\frac{tf_{it}}{Max_{j=1,2,3\ldots,L} tf_{ik}})\log\frac{N}{df_t}}{Max_{j=1,2,3\ldots,L}\{(0.5+0.5\,\frac{tf_{it}}{Max_{j=1,2,3\ldots,L} tf_{ik}})\log\frac{N}{df_j}\}} \quad (2)$$

Where, tf_{it} is the frequency of term t appearing in document d_i, df_t is the total number of documents containing the term t . L is the number of terms contained by d_i, and N is the number of documents in the corpus.

If-Then rule 1: If vector v_i and vector v_j include almost identical words and the weights of the words in both vectors are equivalent, then these two vectors should be similar and have an IVF association relationship. From fig. 5, it can illustrate both (स्कूटर, मोटरसाइकल) has almost the same terms with equivalent weights (दो पहिया , ब्रेक, तेज, गति, धातु). So, there is IVF Ãs relationship with membership degree interval (स्कूटर, मोटरसाइकल [0.8, 0.9]).



Fig. 5. Vector for words (स्कूटर, मोटरसाइकल)

Just Accepted

4.2.2 Identification of IVF Hypernymy (हि) and Hyponymy (हि) relation is done using two methods

1) indicative lexico-syntactic patterns, i.e. particular pattern indicates a semantic relation between terms. In literature various authors took an interest in the automatic acquisition of lexical relations with the goal of building up large lexicons for natural language processing [53,54]. In the section, we also discovered a set of lexico-syntactic patterns to identify IVF hypernymy/hyponymy relation from the text. Below is the list of discovered lexico-syntactic rules, followed by illustrative sentences.

Rule 1: NP_0 जैसे {NP_1, NP_2,….,(and/or)} NP_n}

such that it implies

For all NP_i, $1 \leq i \leq n$, IVF hyponymy (NP_i, NP_0)

For instance: *S1*: कई प्रकार के मिठाई जैसे शकरपारे, आइस्क्रीम, गुड ,Thus from *S1* we conclude: IVF hyponym (“शकरपारे”, “मिठाई” [0.6,0.7]), IVF hyponym (“आइस्क्रीम”, “मिठाई” [0.7, 0.8]), IVF hyponym (“गुड”, “मिठाई” [0.3, 0.4])

Rule 2:NP_0{,NP_i}* और NP_n NP_{n+1}

such that it implies

For all NP_i, $1 \leq i \leq n-1$, IVF hyponymy (NP_i, NP_{n+1}), IVF hyponymy (NP_0, NP_{n+1}), IVF hyponymy (NP_n, NP_{n+1})

For instance: *S2*: कैंची और नेलकटर हथियार है

Thus from *S2* we conclude: IVF hyponym (“कैंची”, “हथियार” [0.4,0.5]), IVF hyponym (“नेलकटर ”, “हथियार”[0.3, 0.4).

Rule 3: NP_0 एक NP_1 है

such that it implies IVF hyponym (NP_1, NP_0)

For instance: *S3*: किशमिश एक फल है

Thus from *S3*, we conclude: IVF hyponym (“किशमिश”, “फल” [0.7,0.8])

Rule 4: NP_0{,NP_i}* (और/ या) बाकी NP_{n+1}

such that it implies For all NP_i, $1 \leq i \leq n-1$, IVF hyponymy (NP_i, NP_{n+1}), IVF hyponymy (NP_0, NP_{n+1}), IVF hyponymy (NP_n, NP_{n+1})

For instance: *S4*:, नील, खरोच, जलना ओर बाकी घाव …..

Thus from *S4* we conclude: IVF hyponym (“नील”, “घाव” [06, 0.7]), IVF hyponym (“खरोच”, “घाव” [0.5,0.6]), IVF hyponym (“जलना”, “घाव”[0.8,0.9])

For instance: *S5*: मीनार, गुफा, बॅवोली या बाकी एतिहसिक स्मारक….

Thus from *S5* we conclude: IVF hyponym (“मीनार”, “एतिहसिक स्मारक”[0.8,0.9]), IVF hyponym (“गुफा”, “एतिहसिक स्मारक”[0.6, 0.7]), IVF hyponym (“बॅवोली”, “एतिहसिक स्मारक” [0.7, 0.82])\

Rule 5: NP_0 विशेष रूप से {NP_1, NP_2,….,(and/or)} NP_n}

such that it implies

For all NP_i, $1 \leq i \leq n$, IVF hyponymy (NP_i, NP_0)

Just Accepted

For instance: *S6*: कॉलेज कर्मचारी विशेष रूप से प्रोफेसर, अनुसंधान साथी

Thus from *S6* we conclude: IVF hyponym ("कॉलेज कर्मचारी", "प्रोफेसर"), IVF hyponym ("कॉलेज कर्मचारी", "अनुसंधान साथी"),

Rule 6: $NP_0\{,NP_i\}^*$ सहित NP_{n+1}

such that it implies For all NP_i, $1 \leq i \leq n-1$, IVF hyponymy (NP_i, NP_{n+1}), IVF hyponymy (NP_0, NP_{n+1}), IVF hyponymy (NP_n, NP_{n+1})

For instance: *S7*: चाकू, सुई, नई कटर सहित सभी हथियार

Thus from *S7* we conclude: IVF hyponym ("हथियार", "नई कटर"), IVF hyponym ("हथियार", "सुई")

2) Creating vectors (v_i, v_j) of terms as discussed above and assigning them weight using tf-idf,

If-then rule 2: if $v_i \subset v_j$ and $W_t(v_i) > W_t(v_j)$, then vector v_i is said to dominate vector v_j and should be more specific than vector v_j

If one vector v_i is a subset of another vector v_j and the weight of terms in v_i vector is larger than those in vector v_j then there is IVF hyponymy/ hypernymy relation exists. An instance is shown in Fig. 6.



Fig. 6. Vector for words (लड़की, औरत)

4.2.3 Identification of IVF Meronymy (M) and Holonymy (Ho) Relation:

1) identified using lexico-syntactic text patterns, i.e. particular pattern indicates a semantic relation between terms.

Rule 1:NP_0 मे NP_1 such that it implies IVF meronymy (NP_1, NP_0)

For instance: *S8*: लॅपटॉप मे बॅटरी होती है , Thus from *S8* we conclude: IVF meronymy ("बॅटरी", "लॅपटॉप" [0.8,0.9])

Rule 2: NP_0, NP_1 हिस्सा such that it implies IVF meronymy (NP_0, NP_1)

For instance: *S9*: बाजू कमीज़ का हिस्सा है , Thus from *S9* we conclude: IVF meronymy ("बाजू", "कमीज़" [0.6,0.7])

Rule 3: NP_0, NP_1 (ओर NP_2)* शामिल है such that it implies IVF meronymy (NP_1, NP_0) and IVF meronymy (NP_2, NP_0)

For instance: *S10*: टीम मे खिलाड़ी ओर सहायक शामिल है, Thus from *S10* we conclude: IVF meronymy ("खिलाड़ी", "टीम") and IVF meronymy ("सहायक", "टीम" [0.5, 0.7])

Rule 4: NP_0, NP_1 दिखाई देते है such that it implies IVF meronymy (NP_0, NP_1)

For instance: *S11*: तारे आसमान मे दिखाई देते है , Thus from *S11* we conclude: IVF meronymy ("तारे", "आसमान" [0.9, 0.98])

Rule 5: NP_0 पर, NP_1 है such that it implies IVF meronymy (NP_1, NP_0) For instance: *S12*: कॅमरा पर धूल है, Thus from *S12*, we conclude: IVF meronymy ("धूल", "कॅमरा" [0.6, 0.7)

Rule 6: NP_0, NP_1 पर उपलब्ध है such that it implies IVF meronymy (NP_0, NP_1)

For instance: *S13*: ग्लोकोज़ दुकान पर उपलब्ध है, thus from *S13* we conclude: IVF meronymy ("ग्लोकोज़", "दुकान" [0.6, 0.8])

Just Accepted

Rule 7: NP_0, NP_1 पाया such that it implies IVF meronymy (NP_1, NP_0)

For instance: *S14*: ग्रह पर पानी पाया, thus from *S14* we conclude: IVF meronymy ("पानी", "ग्रह" [0.5, 0.6])

2) Creating vectors above (v_i, v_j) of terms as discussed above and assigning them weight using tf-idf (An instance shown in Fig. 7).

If-then rule 3: If $v_i \cap v_k = \varnothing$ but $v_i \cup v_k$ is a partition of another vector v_j then v_i and v_j are in a part-of relationship with v_k .

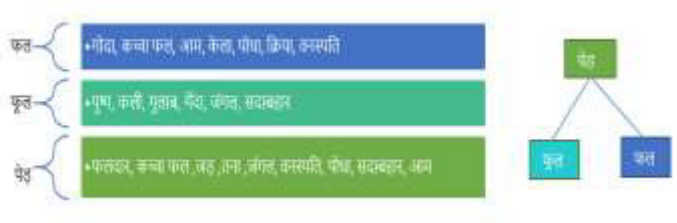


Fig. 7. Vector for words (फल, फूल, पेड़)

4.2.4 *Identification of IVF antonyms (Ã) Relation:*

Identification of IVF antonymy relation depends on fuzzy association relation and crisp antonymy relation.

If-Then rule : If $(\mu_{\tilde{A}s}\ (a, b) = [x, y]$ where $0 \le x, y \le 1) \cap (\mu_A\ (b, c) = 1)$ then $(\mu_{\tilde{A}}\ (a, c) = [x, y]$ where $0 \le x, y \le 1)$. An instance for (औरत, लड़का) is shown in Fig. 8. Where, *a* is "औरत", *b* is "लड़की" and *c* is "लड़का" and $\mu_{\tilde{A}}$ is IVF antonymy, A is crisp antonymy and $\mu_{\tilde{A}s}$ is IVF association relation

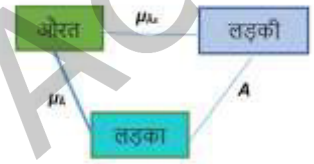


Fig. 8. An instance for Identification of IVF antonymy relation

4.2.5 *Identification of IVF entailment Relation:*

To identify IVF entailment relationships, ConceptNet is Used. The relations "HasSubvent", "HasFirstSubevent", "HasLastSubevent", "HasPrerequisite" helped to identify entailment. Their definition with example is given in the Fig. 9.

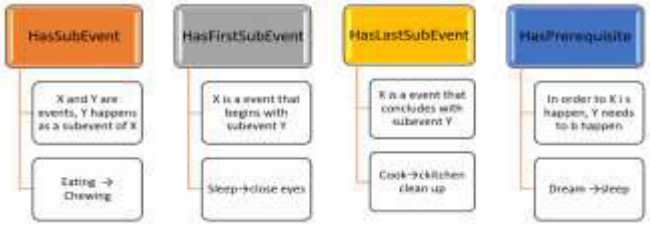


Fig. 9: ConceptNet Relations for Entailment Identification

Another way is to identify is again Lexico-Syntactic Patterns like:

Rule 1: VP_0 मे VP_1 such that it implies IVF meronymy (VP_1, VP_0)

For instance: *S8*: खाना बनाने मे गर्मी लगती है, thus from *S8* we conclude: IVF meronymy ("खाना बनाना", "गर्मी लगना")

Just Accepted

Rule 2: VP_0 से VP_1 such that it implies IVF meronymy (VP_1, VP_0)

For instance: *S10*: मेहनत करने से फल मिलता है, Thus from *S10* we conclude: IVF meronymy ("मेहनत करना", "फल मिलना"). *S11*: व्यायाम करने से स्वस्थ रहते है, Thus from *S11* we conclude: IVF meronymy ("व्यायाम करना", "स्वस्थ रहना")

Rule 3: VP_0 के बाद VP_1 such that it implies IVF meronymy (VP_1, VP_0)

For instance: *S12*: सोने के बाद खराटे आते है, Thus from *S12* we conclude: IVF meronymy ("सोना", "खराटे आना").

4.2.6 Identification of IVF Troponymy Relation:

IVF troponymy relation exists between two verbs while IVF hypernymy/hyponymy exists between two verbs. So lexico-syntactic patterns and If-Then rules used for IVF hypernymy/hyponymy relation can be used for identification of IVF troponymy. To identify IVF troponymy relationships, ConceptNet is used. The relations "MannerOf", helped to identify entailment. Its definition, with the example, is given in Fig. 10.

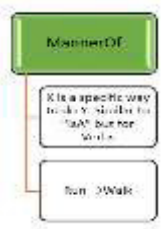


Fig.10: ConceptNet Relation for Troponymy Identification

4.2.7 Identification of IVF Ability, Capability and Function Link Relation:

The ability relation is used to specify the inherited characteristics of a nominal concept. Capability relation specifies the acquired characteristics of a nominal term and last function relation specify the functions of a nominal concept. These relations are learned from the data. All the NP+VP can be extracted for this relationship

4.2.8 Identification of IVF Attribute Relation:

This relation is learned from the data. All the NP+Adj can be extracted for this relationship. IVF attributes specify the partial property of a noun that can be only extracted using noun + adjective phrases from text.

4.2.9 Identification of Causative, Modifies Noun, Modifies Verb, Derived Form Relation:

These relations are always crisp relations with membership value unity. So, there is no need to identify them; specifically, these relationships can be extracted from Hindi WordNet directly.

4.3 Automatic Assignment of Membership Degree to Interval-Valued Fuzzy Relations

In literature, the multiple experts' opinion has been used to assign membership value to relations between two words/nodes in fuzzy graph [4]. But it is a tedious and challenging task to determine the exact membership value by experts between millions of vertices/words. In this section, the automatic assignment of membership value to edges of a fuzzy graph is done by using word embeddings. In the proposed method, we used various word embeddings, including Word2Vec, Glove, Fasttext and BERT to find the belongingness between two words in a fuzzy graph. Assignment of Association, Hypernymy, Hyponymy, Antonymy, Meronymy, Holonymy, Entailment, Troponymy, Ability, Capability, Function Attribute relations are calculated using embeddings. And causative, Modifies Noun, Modifies Verb, Derived Form relation are always a crisp relation with membership value unity. For the dataset, we crawled a huge monolingual Hindi corpus from Hindi Wikipedia of size 7.2 GB. Words with a frequency of less than 2 in the total corpus are classified as unknown (out-of-vocabulary) terms. After preprocessing word2vec embeddings are calculated using the cosine similarity formula as follows:

Similarity score: It is used to denote the belongingness between the two nodes, and it's value lies between 0 and 1. The angle between two vectors is employed as a measure of divergence between the vectors, and the cosine of this angle is used to calculate cosine similarity. It is calculated using Equation (3) [40].

Just Accepted

$$similarity(i, j) = \frac{i.j}{|i| \times |j|} \quad (3)$$

where, *i.j* is the dot product of the two vectors and is given by Equation (4) and *|i|, |j|* represents the magnitude of the two vectors and is given by Equation (5a)-(5b) and *similarity(i,j)* represents the cosine similarity between the two vectors, where n is the size of the vector.

$$i.j = \sum_{k=0}^{n} i_k \times j_k \quad (4)$$

$$|i| = \sqrt{\sum_{k=0}^{n} i_k^2} \quad (5a)$$

$$|j| = \sqrt{\sum_{k=0}^{n} j_k^2} \quad (5b)$$

Next to build FastText embeddings , the gensim library was used for implementation of FastText with dimensions of 50, 100, 200, and 300 (skip-gram architecture) [20,41]. once again. The default settings of gensim are utilised for the remaining parameters. FastText's official repository contains pre-trained word embeddings for the Hindi language. For GloVe embeddings [19] of 50, 100, 200, and 300 dimensions are created. A symmetric window of size 15 is used to build the co-occurrence matrix.

We train 300-dimensional BERT (Bidirectional Encoder Representations from Transformers) embeddings [21]. We employ sentence component embeddings [40], which were trained on a corpus with a vocabulary of 25000 words. This model was pre-trained using three Tesla K80 GPUs, each with 12 GB of memory. The official BERT repository has a multilingual model with 102 languages, including Hindi.

Let values provided by these four embeddings are: W_{2v}, G_L, F_T, and B_E

$$\mu_r = [min\ (W_{2v}, G_L, F_T, B_E),\ max\ (W_{2v}, G_L, F_T, B_E)] \quad (6)$$

where μ_r is the IVF membership value of the relation *r*.

5 EXPERIMENTAL SETUP AND RESULT DISCUSSION

The proposed Interval- Valued Fuzzy Hindi WordNet is evaluated using a heath domain corpus, publicly available at the IIT Bombay website [8]. Health corpus is part-of-speech tagged and comprises 90 files with approximately one hundred sentences in each file. In the corpus, the average number of open-class terms in a sentence is 8.85 ≈ 9. The number of nouns, Verbs and adjectives required for different IVF relations (mention in section) is 52230, 24291 and 22699, respectively.

Table I shows some of the sentences from the Hindi corpus, as well as the fuzzy relationships between the terms in the sentences. The table also mentions membership value taken from linguistic experts and membership value computed by the proposed method. The mean square error (MSE) between the expert value and the proposed value is calculated to evaluate the proposed approach.

MSE calculates that how much error the method has made during prediction [52]. The MSE between the EV (Expert Value) and the value calculated by the proposed method (PM) can be calculated as:

$$MSE\ (EV, PM) = \frac{1}{n} \sum_{i=1}^{n} [EV_i - PM_i]^2 \quad (7)$$

Just Accepted

To obtain expert value, we enlisted the help of a group of recognised Hindi lexicographers who are experts in linguistics, the Hindi language, and fuzzy logic. Every expert in the room assigned a value between 0 to 1 to the relationship between each pair of words. An interval value expressing the strength between words is examined, according to the experts' own knowledge and perception. As we know, it is impossible to get values for billions of relations from experts, so we randomly selected 60 relations between nouns, 50 relations between verbs, 20 relations between nouns & verbs and 20 relations between nouns & adjectives. Finally, MSE for relations between nouns, between verbs, between noun & verbs and nouns & adjectives are shown in Table II and their graphical representation is shown in Fig. 11. The results demonstrated for the selective words pair proposed method (IVFHWN) is providing good results. The proposed method can make the concept of fuzzifying Hindi WordNet feasible for all applications of natural language processing.

In literature, classical Hindi WordNet is widely used for Word Sense Disambiguation (WSD) [44, 4] so, we applied the proposed concept of IVFHWN to perform WSD at heath domain tagged dataset by IIT Bombay [8]. To achieve WSD, centrality measures including fuzzy local and global graph connectivity measures, are calculated on the WordNet graph. These fuzzy local and global graph connectivity measures are proposed by Jain at. al. and used by many authors [4, 16, 45, 46]. We applied type reduction and defuzzification on the computed intervals for the relation between two Hindi words. To evaluate the proposed system, we used recall, precision and their harmonic mean F-measure matrices. Since the proposed method offers an answer for all ambiguous inputs so in our case recall, precision and f-measure are the same.

For the comparison, we used three lexicons: Hindi WordNet, Fuzzy Hindi WordNet (where membership values are taken from linguistic experts) and Proposed IVFHWN (where membership values are automatically computed using word embeddings and interval value fuzzy logic). The comparison results on a subset of the health corpus are shown in Table III. In this table, the column “All” and column “poly” shows the result on all words (polysemous and monosemous) and results only on the subset of polysemous words, respectively. Graphical representation of comparison is shown in Fig. 12. As can be seen in the results, Fuzzy Hindi WordNet outperforms Classical Hindi WordNet. This is because Fuzzy Hindi WordNet handles the real-life uncertainties and understands the words better while performing WSD. Next, it is also visible that fuzzy Hindi WordNet performs even better than the proposed IVFHW. The reason for this is the membership values in Hindi fuzzy WordNet are taken from various linguistic experts. In IVFHW, we automatically computed the membership values. The difference between the results is not very high, and the proposed system provides the ability to fuzzy a large number of words in a corpus which is not possible in the case of the involvement of linguistic experts.

In the proposed IVFHW, out of local and global measures, local measures are providing better results. In local measure, Degree and page rank are proving better results than the remaining three. In global edge density and compactness measures are proving better results than graph entropy.

Just Accepted

TABLE I Sample sentences from the Hindi corpus with the fuzzy relationships between the terms

	Hindi Sentences	Pre-Processes Open class words	Identified Fuzzy Relationship between open class words
1	शतरंज, सुडोकू पहेलिया आदि जैसे खेल खेले	शतरंज, सुडोकू पहेलिया. खेल	पहेलिया Fuzzy Hypernymy खेल
2	अपने विचारो को डायरी मे लिख ले	विचार, डायरी	विचार Fuzzy Meronymy डायरी
3	दिमाग़ शक्ति को बड़वा देने के लिए दवाईया जैसे ब्रह्मी, अश्वगंधा और कैलामस लें	दिमाग़, शक्ति, दवाईया, ब्रह्मी, अश्वगंधा कैलामस,	दवाईया Fuzzy Hypernymy ब्रह्मी, अश्वगंधा कैलामस
4	गहरी साँस लेने का अभ्यास करें	गहरी, साँस	गहरी, Fuzzy Attribute साँस

TABLE II MEAN SQUARE ERROR (MSE) FOR RELATIONS BETWEEN NOUNS, BETWEEN VERBS, BETWEEN NOUN & VERBS AND NOUNS & ADJECTIVES

	Word	Word	Total number of relations	#Relations considered	MSE
Relation Between Nouns	No. of Noun (N) in dataset: 52230	No. of Noun (N) in dataset: 52230	N×N=2.72 Billion	60	0.63×10^{-2}
Relation Between verbs	No. of Verbs (V) in dataset: 24291	No. of Verbs (V) in dataset: 24291	V×V = 0.59 Billion	50	0.136×10^{-2}
Relation Between Noun and Verbs	No. of Noun (N) in dataset: 52230	No. of Verbs (V) in dataset: 24291	N×V=1.26 Billion	20	0.146×10^{-2}
Relation between Noun and Adjectives	No. of Noun (N) in dataset: 52230	No. of adjectives (Adj) in dataset: 22699	N×Adj= 1.18 Billion	20	0.13×10^{-2}



Fig. 11. Plots of Mean Square Error (MSE) for relations between nouns, between verbs, between noun & verbs and nouns & adjective

TABLE III CONNECTIVITY MEASURES PERFORMANCE (F1 MEASURES)

		Classical Hindi WordNet [1]		Fuzzy Hindi WordNet/ Value taken from linguistics Expert [4]		Proposed IVFHW	
		All	Poly	All	Poly	All	Poly
Local	Fuzzy Degree	0.53	0.41	0.62	0.49	0.61	0.48
	Fuzzy PageRank	0.53	0.40	0.61	0.49	0.58	0.48
	Fuzzy HITS	0.45	0.32	0.53	0.44	0.47	0.41
	Fuzzy KPP	0.47	0.37	0.56	0.45	0.49	0.44
	Fuzzy Betweenness	0.49	0.37	0.59	0.49	0.57	0.46
Global	Fuzzy Compactness	0.42	0.30	0.49	0.40	0.43	0.39
	Fuzzy Entropy	0.41	0.28	0.45	0.31	0.37	0.29
	Fuzzy Edge Density	0.43	0.30	0.51	0.42	0.46	0.35

Just Accepted

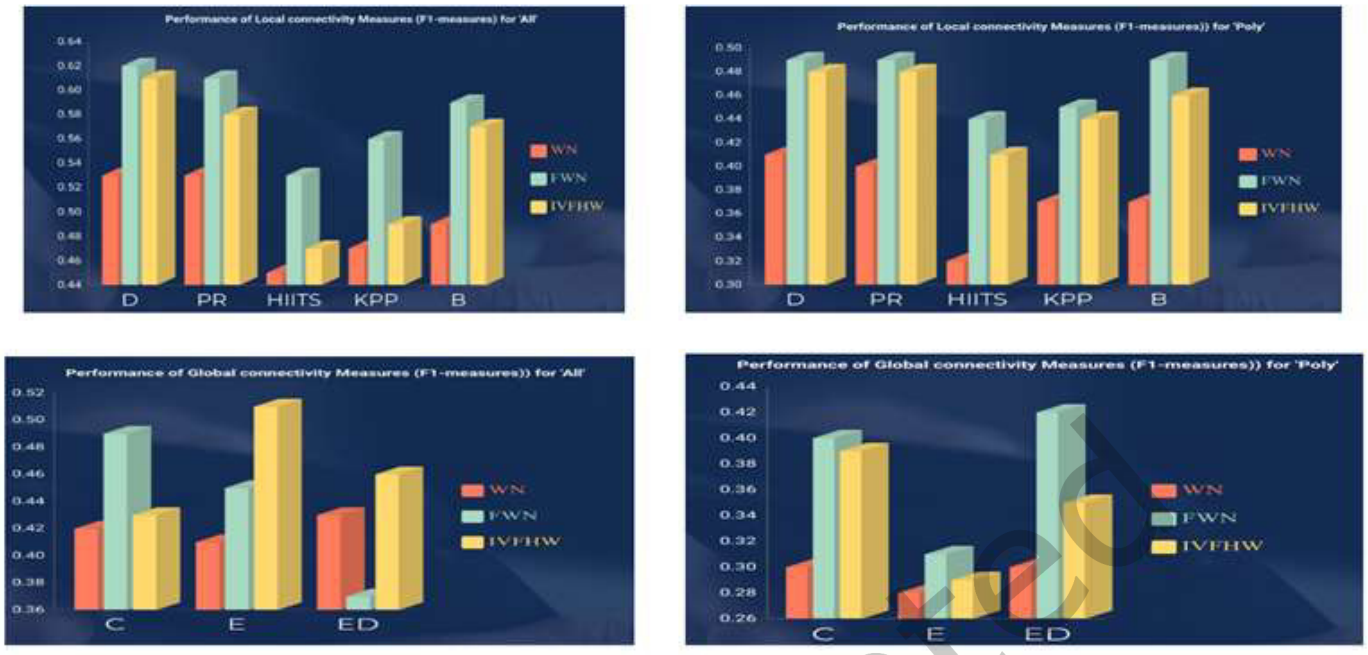


Fig. 12. Graphical representation of performance of local and connectivity measures for "All" and "Poly" using Hindi WordNet, Fuzzy WordNet and proposed IVFHWN

6 CONCLUSION AND FUTURE WORK

Hindi WordNet is an extremely useful resource for dealing with Natural Language Processing challenges. In literature, it is confirmed that fuzzification of Hindi WordNet shows better results than classical Hindi WordNet. But in existing work, membership values for relations are taken from experts, which is not feasible in millions of words. This paper introduces the automatic assignment of membership values to the relations between two Hindi words using lexico-syntactic patterns and various word embeddings. This automation reduces the expert's requirement and makes it feasible for any number of words. This work proposes the Interval- Valued Fuzzy relations for Hindi WordNet, which handles the uncertainty of type 1 fuzzy Hindi WordNet. For the experimentation, the proposed Interval- Valued Fuzzy Hindi WordNet is evaluated using a heath domain corpus and The results verified that (IVFHWN) is providing good results. The proposed concept is also compared with classical HWN, Fuzzy Hindi word for word sense disambiguation problem of NLP. We found IVFHWN gives better results than classical HW and FHWN. Since WordNet exists for several languages, for instance IndoWordNet (for Indian languages), EuroWordNet etc., establishing and evaluating the performance of extending WordNet to Interval-Valued WordNet with respect to these resources is an intriguing future direction. For many lexicons such as sentiwordnet, WordNet is used as a based, so extension of sentiwordnet to Interval- Valued Fuzzy sentiwordnet can improve resource performance. It would be interesting to see how much accuy improved by using IVFHN In the future IVFHW can be used for the other purposes of NLP, including Information keyphrase extraction, Text summarisation, query expansion etc.. To handle uncernity in more realistic way (more cose to rea world)The incorporation of type-3 fuzzy logic in Hindi WordNet can be explored in future.

Just Accepted

APPENDIX 1

Hindi Word	Transliteration	Corresponding English	Hindi Word	Transliteration	Corresponding English
सोफा	*sophā*	Sofa	मतदान	*matadāna*	Vote
दस्ता	*dastā*	A quire of Paper	इंसान	*iṃsāna*	Human
कॉपी	*kapī*	Copy	पिलाना	*pilānā*	To feed
मेज	*meja*	Table	नौकरी करना	*Naukarī karanā*	To do job
स्टूल	*sṭūla*	Stool	लड़का	*laḍakā*	Boy
नेल कटर	*nela kaṭara*	Nail Cutter	आदमी	*ādamī*	Man
हथियार	*hathiyāra*	Weapon	पंछी	*paṃchī*	Bird
बईमान	*bīmāna*	Dishonest	उड़ना	*uḍanā*	To Fly
दुर्जन	*durjana*	Wicked	पढ़ाना	*paḍhaṭānā*	To Study
बटन,	*baṭana*	Button	संस्कारी	*saṃskārī*	Cultured
कुर्ता	*kurtā*	Shirt	सुंदर	*suṃdara*	Beautiful
फूल	*phūla*	Flower	सीखना	*sīkhanā*	To Learn
पेड़	*peḍa*	Tree	भागना	*bhāganā*	To Run
बिजली	*bijalī*	Electricity	व्यायाम करना	*Vyāyāma karanā*	Do exercise
बारिश	*bāriśa*	Rain	बेहोश	*behośa*	Faint
बेटी	*beṭī*	Daughter	शाम	*śāma*	Evening
मुर्दा	*murdā*	Dead	सिलाई करना	*Silāī karanā*	To do stitching
खुश होना	*khuśa honā*	To Smile	चलना	*calanā*	Walk
बुरी आदते होना	*burī ādate honā*	Having bad Habits	निरोगी होना	*Nirogī honā*	To be healthy
खाना बनाना	*khānā banānā*	To cook	स्कूटर	*skūṭara*	Scooter
कुर्सी	*kursi*	Chair	तेज	*teja*	Fast
राय	*rāya*	Opinion	फल	*phala*	Fruit
पीना	*pīnā*	To Drink	बंजर	*baṃjara*	Barren
सोना	*sonā*	To Sleep	भूमि	*bhūmi*	Earth
सुलाना	*sulānā*	Put to sleep	मोटरसाइकल	*moṭarasāikala*	Motorcycle
अध्यापक	*adhyāpaka*	Teacher	शकरपारे	*śakarapāre*	Shakarpare
सुशील	*suśīla*	Gentle	आइस्क्रीम	*āiskrīma*	Ice-cream
लड़की	*laḍaṭkī*	Girl	मिठाई	*miṭhāī*	Sweets
बुरे संस्कार होना	*bure saṃskāra honā*	To have bad moral	गुड	*guḍa*	Jaggery
सब्जी काटना	*sabjī kāṭanā*	Vegetable cutting	दिन	*dina*	Day
सुंदरता	*suṃdaratā*	Beauty	हसना	*hasanā*	To laugh
सुंदरी	*suṃdarī*	Beauty			

Hindi Sentences	Transliteration	Corresponding English
कैंची और नेलकटर हथियार है	*keṁcī aura nelakaṭara hathiyāra hai*	Scissors and nail cutters are weapons
किशमिश एक फल है	*kiśamiśa eka phala hai*	Raisin is a fruit
नील, खरोच, जलना ओर बाकी घाव	*nīla, kharoca, jalanā ora bākī ghāva*	bruises, bruises, burns and other wounds
मीनार, गुफा, बॅवोली या बाकी एतिहसिक स्मारक	*mīnāra, guphā, baॅvolī yā bākī etihasika smāraka*	minaret, cave, baoli or other historical monument
अधिकांश कॉलेज कर्मचारी विशेष रूप से प्रोफेसर, शिक्षक सह अनुसंधान साथी	*adhikāṃśa kaॉlejeja karmacārī viśeṣa rūpa se prophesara, śikṣaka saha anusaṃdhāna sāthī*	Most of the college staff especially professors, teachers cum research fellows
चाकू, सुई, नई कटर सहित सभी हथियार	*cākū, suī, nī kaṭara sahita sabhī hathiyāra*	All weapons including knives, needles, new cutters
लॅपटॉप मे बॅटरी होती है	*laॅpaṭaॉpa me baॅṭarī hotī hai*	laptop has battery
बाजू कमीज़ का हिस्सा है	*bājū kamīja़ kā hissā hai*	the sleeve is part of the shirt
टीम मे खिलाड़ी ओर सहायक शामिल है	*ṭīma me khilāḍaṭī ora sahāyaka śāmila hai*	Team includes players and assistants
तारे आसमान मे दिखाई देते है	*tāre āsamāna me dikhāī dete hai*	stars appear in the sky
ग्लोकोज़ दुकान पर उपलब्ध है	*glokoja dukāna para upalabdha hai*	Available at Glucose Shop
ग्रह पर पानी पाया	*graha para pānī pāyā*	found water on the planet
खाना बनाने मे गर्मी लगती है	*khānā banāne me garmī lagatī hai*	Feel hot while cooking

Just Accepted

मेहनत करने से फल मिलता है	*mehanata karane se phala milatā hai*	hard work pays off
व्यायाम करने से स्वस्थ रहते है	*vyāyāma karane se svastha rahate hai*	stay healthy by exercising
सोने के बाद खराटे आते है	*sone ke bāda kharāṭe āte hai*	Sore comes after sleeping
शतरंज, सुडोकू पहेलिया आदि जैसे खेल खेले	*śataraṃja, suḍokū paheliyā ādi jaise khela khele*	Play games like chess, sudoku puzzles etc.
अपने विचारो को डायरी मे लिख ले	*apane vicāro ko ḍāyarī me likha le*	write your thoughts in a diary
दिमाग़ शक्ति को बड़वा देने के लिए दवाईया जैसे ब्रह्मी, अश्वगंधा और कैलामस लें	*dimāga śakti ko baḍavā dene ke lie davāīyā jaise brahmī, aśvagaṃdhā aura kailāmasa leṃ*	To increase brain power, take medicines like Brahmi, Ashwagandha and Calamus
गहरी साँस लेने का अभ्यास करें	*gaharī sāṁsa lene kā abhyāsa kareṃ*	practice deep breathing

REFERENCES

[1] G. A. Miller. 1995. WordNet: A Lexical Database for English. *Communications of the ACM* Vol. 38, No. 11: 39-41.

[2] D. Narayan, D. Chakrabarti, P. Pande, and P. Bhattacharyya, 2002. An experience in building the indo WordNet-a WordNet for Hindi. *First International Conference on Global WordNet, Mysore, India.* Vol. 24.

[3] A. Jain, M. Jain, G. Jain, and D. K. Tayal, 2019. 'UTTAM': An efficient spelling correction system for Hindi language based on supervised learning, *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 18, no. 1, pp. 1–26, , doi: 10.1145/3264620.

[4] A. Jain and D. K. Lobiyal, 2015.Fuzzy Hindi wordnet and word sense disambiguation using fuzzy graph connectivity measures, *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 15, no. 2, doi: 10.1145/2790079.

[5] S. K. Ray, A. Ahmad, and K. Shaalan, 2018. A review of the state of the art in hindi question answering systems," *Stud. Comput. Intell.*, vol. 740, pp. 265–292, doi: 10.1007/978-3-319-67056-0_14.

[6] P. Sharma and N. Joshi, 2019.Knowledge-Based Method for Word Sense Disambiguation by Using Hindi WordNet, *Eng. Technol. Appl. Sci. Res.*, vol. 9, no. 2, pp. 3985–3989, doi: 10.48084/etasr.2596.

[7] A. Das and S. Bandyopadhyay, 2010. SentiWordNet for Indian Languages, *8th Work. Asian Lang. Resour. (ALR), August*, no. August, pp. 56–63, [Online]. Available: http://www.aclweb.org/anthology/W/W10/W10-3208.pdf.

[8] Hindi Corpus from Center for Indian Language Technology Solutions, IIT Bombay, Mumbai, India http://www.cfilt.iitb.ac.in/Downloads.html

[9] J. A Goguen, LA Zadeh. Fuzzy sets. Information and control, vol. 8, pp. 338–353.

[10] J. C. Bezdek, G. Biswas, and L. ya Huang, 1986.Transitive closures of fuzzy thesauri for information-retrieval systems, *Int. J. Man. Mach. Stud.*, vol. 25, no. 3, pp. 343–356, 1986, doi: 10.1016/S0020-7373(86)80065-7.

[11] S. M. Chen, Y. J. Horng, C. H. Lee, 2001. Document Retrieval Using Fuzzy Valued Concept Network. *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 31, No. 1, pp.111-18.

[12] S. M. Chen, Y. J. Horng, and C. H. Lee 2003. , Fuzzy information retrieval based on multi-relationship fuzzy concept networks, *Fuzzy Sets Syst.*, vol. 140, no. 1, pp. 183–205, doi: 10.1016/S0165-0114(02)00464-5.

[13] M. De Cock, S. Guadarrama, and M. Nikravesh,2005. Fuzzy Thesauri for and from the WWW, pp. 275–284, doi: 10.1007/3-540-32365-1_8.

[14] C. Diou, G. Katsikatsos, and A. Delopoulos,2006. Constructing fuzzy relations from WordNet for word sense disambiguation, *Proc. - SMAP 2006 1st Int. Work. Semant. Media Adapt. Pers.*, pp. 135–140, doi: 10.1109/SMAP.2006.14.

[15] M. Kracker, 1992. A fuzzy concept network model and its applications, pp. 761–768, doi: 10.1109/fuzzy.1992.258752.

[16] A. Jain, S. Vij, and O. Castillo, 2019. Hindi query expansion based on semantic importance of Hindi WordNet relations and fuzzy graph connectivity measures, *Comput. y Sist.*, vol. 23, no. 4, pp. 1337–1355, doi: 10.13053/CyS-23-4-2984.

[17] Y. Bengio, R. Ducharme, P. V. 2003. A neural probabilistic language model. *In Journal of Machine Learning Research*, pp 3:1137–1155.

[18] T. Mikolov, K. Chen, G. Corrado, & J. Dean, 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781.*

[19] J. Pennington, R. Socher, and C. Manning, 2014, Glove: Global vectors for word representation. *In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp 1532– 1543.

[20] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, 2017.Enriching Word Vectors with Subword Information, *Trans. Assoc. Comput. Linguist.*, vol. 5, no. June, pp. 135–146, doi: 10.1162/tacl_a_00051.

[21] J. Devlin, M.W. Chang, K. Lee, & K. Toutanova, 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[22] A. K. Pujari, A. Mittal, A. Padhi, A. Jain, M. Jadon, and V. Kumar, 2019.Debiasing gender biased hindi words with word-embedding, *ACM Int. Conf. Proceeding Ser.*, pp. 450–456, doi: 10.1145/3377713.3377792.

[23] S. Srigiri, & S. K. Saha, 2018. Spelling Correction of OCR-Generated Hindi Text Using Word Embedding and Levenshtein Distance. In *International Conference on Nanoelectronics, Circuits and Communication Systems* (pp. 415-424). Springer, Singapore.

[24] A. Gupta, R. Sukumaran, K. John, and S. Teki, 2021. Hostility Detection and Covid-19 Fake News Detection in Social Media,. Available: http://arxiv.org/abs/2101.05953.

[25] T. S. Aparna, K. Simran, B. Premjith, & K. P. Soman, 2021. Aspect-Based Sentiment Analysis in Hindi: Comparison of Machine/Deep Learning Algorithms. In *Inventive Computation and Information Technologies* (pp. 81-91). Springer, Singapore.

[26] B. Shah and S. K. Kopparapu, 2019. A Deep Learning approach for Hindi Named Entity Recognition. Available: http://arxiv.org/abs/1911.01421.

[27] A. Kumari and D. K. Lobiyal, 2021.Efficient estimation of Hindi WSD with distributed word representation in vector space, *J. King Saud Univ. - Comput. Inf. Sci.*, doi: 10.1016/j.jksuci.2021.03.008.

[28] S. Bhingardive, R. Puduppully, D. Singh, and P. Bhattacharyya, 2014. Merging Verb Senses of Hindi WordNet using Word Embeddings, *Proc. 11th Int. Conf. Nat. Lang. Process.*, no. December, pp. 344–352. Available: https://www.aclweb.org/anthology/W14-5148.

[29] R. Kumar, P. Jha, and V. Sahula, 2019.An augmented translation technique for low resource language pair: Sanskrit to hindi translation, *ACM Int. Conf. Proceeding Ser.*, pp. 377–383, doi: 10.1145/3377713.3377774.

[30] A. Jain, S. Yadav, and D. Tayal, 2013.Measuring context-meaning for open class words in Hindi language, *2013 6th Int. Conf. Contemp. Comput. IC3 2013*, no. May 2016, pp. 118–123, doi: 10.1109/IC3.2013.6612174.

[31] M. Sinha, M. Kumar, P. Pande, L. Kashyap, and P. Bhattacharyya. 2004, November. Hindi word sense disambiguation. In International Symposium on Machine Translation, Natural Language Processing and Translation Support Systems, Delhi, India.

[32] S. K. Dwivedi and P. Rastogi, 2008.An entropy based method for removing web query ambiguity in Hindi language, *J. Comput. Sci.*, vol. 4, no. 9, pp. 762–767, doi: 10.3844/jcssp.2008.762.767.

[33] S. Das, A. Seetha, M. Kumar, & J. L. Rana, 2010. Post translation query expansion using Hindi Word-Net for English-Hindi CLIR system. In *Proceedings of the Forum for Information Retrieval Evaluation (FIRE'10).*

[34] A. Sharan, M. L. Joshi, & A. Pandey, 2011. Exploiting ontology for concept based information retrieval. In *International Conference on Information Systems for Indian Languages* (pp. 157-164). Springer, Berlin, Heidelberg.

[35] A. Jain, D. K. Tayal, and S. Rai, 2015. Shrinking digital gap through automatic generation of WordNet for Indian languages, *AI Soc.*, vol. 30, no. 2, pp. 215–222, doi: 10.1007/s00146-014-0548-5.

[36] A. Jain, D. K. Tayal, and S. Yadav, 2016.Retrieving web search results using Max–Max soft clustering for Hindi query, *Int. J. Syst. Assur. Eng. Manag.*, vol. 7, no. December, pp.

Just Accepted

70–81, doi: 10.1007/s13198-014-0307-5.
[37] C. Cornelis, G. Deschrijver, and E. E. Kerre, 2004. Implication in intuitionistic fuzzy and interval-valued fuzzy set theory: construction, classification, application. *International journal of approximate reasoning*, *35*(1), pp.55-95.
[38] M. Akram, & W. A. Dudek, 2011. Interval-valued fuzzy graphs. *Computers & Mathematics with Applications*, *61*(2), 289-299.
[39] S. Kumar, V. Mansotra, and R. Goswami. 2013. Query Optimisation: Solution for low recall problem in Hindi Language IR--Revisited with Experimental Results and Analysis. *Language in India*, *13*(1).
[40] Google. 2018. Sentence piece embeddings. https:// github.com/google/sentencepiece.
[41] R. Rehurek and P. Sojka, 2010. Software Framework for Topic Modelling with Large Corpora, *Proc. Lr. 2010 Work. New Challenges NLP Fram.*, pp. 45–50.
[42] R. Speer, J. Chin, and C. Havasi, 2016. ConceptNet 5.5: An Open Multilingual Graph of General Knowledge, no. Singh 2002, pp. 4444–4451, Available: http://arxiv.org/abs/1612.03975.
[43] H. Rashmanlou, and Y. B. Jun, 2013. Complete interval-valued fuzzy graphs. *Annals of Fuzzy Mathematics and Informatics*, *6*(3), pp.677-687.
[44] R. Navigli and M. Lapata, 2010. An experimental study of graph connectivity for unsupervised word sense disambiguation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 4, pp. 678–692, doi: 10.1109/TPAMI.2009.36.
[45] A. Jain, K. Mittal, and K. S. Vaisla, 2020. FLAKE: Fuzzy Graph Centrality-based Automatic Keyword Extraction, *Comput. J.*, doi: 10.1093/comjnl/bxaa133.
[46] K. Mittal, A. Jain, K. S. Vaisla, and D. K. Tayal. 2020. A Novel method for Reviewer Assignment Problem based on Reviewers' Profile and Fuzzy Graph Connectivity Measure. *In 2020 International Conference on Intelligent Engineering and Management (ICIEM)*, pp. 386-391. IEEE.
[47] Gonçalo Oliveira, H., & P. Gomes, 2011. Automatic Discovery of Fuzzy Synsets from Dictionary Definitions. *22nd International Joint Conference on Artificial Intelligence*, 1801–1806. http://ijcai.org/papers11/Papers/IJCAI11-302.pdf
[48] V. Kann, & M. Rosell, 2005. Free construction of a free Swedish dictionary of synonyms. *Proc. 15th Nordic Conf. on Comp. Ling.–NODALIDA* (5).
[49] S. Vij, A. Jain, D. Tayal, and O. Castillo, 2018. Fuzzy logic for inculcating significance of semantic relations in word sense disambiguation using a WordNet graph. *International journal of fuzzy systems*, *20*(2), pp.444-459.
[50] Y. Alizadeh-Q, B. Minaei-Bidgoli, S. A. Hossayni, M.R. Akbarzadeh-T, D. R. Recupero, M. R. Rajati, and A. Gangemi, 2021. Interval Probabilistic Fuzzy WordNet. *arXiv preprint arXiv:2104.10660*.
[51] A. Joshi, A. R. Balamurali, and P. Bhattacharyya, 2010. A fall-back strategy for sentiment analysis in Hindi: a case study. *Proceedings of the 8th ICON*.
[52] D. M. Allen, 1971. Mean square error of prediction as a criterion for selecting variables. *Technometrics*, *13*(3), pp.469-475.
[53] M. A. Hearst. Automatic acquisition of hyponyms from large text corpora. *In Proceedings of the 14th conference on Computational linguistics*, pages 39–545, Nantes, France, 1992.
[54] C. Klaussner and D. Zhekova, 2011, Lexico-syntactic patterns for automatic ontology building. In *Proceedings of the Second Student Research Workshop associated with RANLP 2011*, pp. 109-114.

Just Accepted

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/377858209

Bond Graph Simulation of Composite Rotor's Nonlinear Response

Chapter · December 2023

DOI: 10.1007/978-981-99-4638-9_34-1

CITATIONS
0

READS
22

5 authors, including:



Shadab Ahmad
Shandong University of Technology
29 PUBLICATIONS **188** CITATIONS
SEE PROFILE



Yebing Tian
Shandong University of Technology
111 PUBLICATIONS **1,150** CITATIONS
SEE PROFILE

All content following this page was uploaded by Shadab Ahmad on 08 February 2024.

The user has requested enhancement of the downloaded file.

Bond Graph Simulation of Composite Rotor's Nonlinear Response

Anuj Sharma, Atul Kumar Agrawal, Shadab Ahmad, Shanay Rab, and Yebing Tian

Contents

Introduction 4
Fabrication and Experimental Impact Hammer Response of Different MMC Shafts 6
Estimation of Effective Mechanical Properties Through Mean-Field Homogenization by Using Parameters Extracted Through Image Processing 8
Bond Graph Modeling 10
Simulation 17
Conclusion 24
References 25

Abstract

In this chapter, an intrinsic damping coefficient is being calibrated by mapping the experimental data with the simulation results, and it is further utilized for dynamic analysis of composite shafts. The experimental impact hammer response of shafts with different compositions of 6061 Al/ Al_2O_3 is being extracted through a digital vibration analyzer OROS36®. Computational results are obtained by a multiscale modeling approach where effective material properties are predicted through the mean-field homogenization. These properties are utilized as parameters for computing the vibration response of these shafts through a multi-energy domain bond graph approach. Microscopic parameters such as the size of particles, average aspect ratio and compositions are determined by image processing of SEM

A. Sharma
Department of Mechanical, Production, Delhi Technological University, Delhi, India

A. K. Agrawal
Industrial and Automobile Engineering, Delhi Technological University, Delhi, India

S. Ahmad (✉) · Y. Tian
School of Mechanical Engineering, Shandong University of Technology, Zibo, China

S. Rab
School of Architecture, Technology and Engineering, University of Brighton, Brighton, UK

© Crown 2024

N. Garg et al. (eds.), *Handbook of Vibroacoustics, Noise and Harshness*,
https://doi.org/10.1007/978-981-99-4638-9_34-1

images of different composite shafts. The experimental investigation clearly depicts the effect of increasing reinforcement in the composite. The amplitudes of vibration are reduced, and the natural frequency of the shafts is increased as the volume percentage of the reinforcement increases. The calibrated parameters provided congruent results with the experimental data.

Keywords

Dynamics and vibration · Multiscale modeling · Bond graph modeling · Composite rotor · Micromechanics · Mean-field homogenization

Nomenclature

N	Number of phases
ϑ_m	Volume fraction of metal matrix
ϑ_i	Volume fraction of reinforcement
$\bar{\bar{K}}_{av}$	Overall elastic stiffness
K_m	Matrix stiffness
K_i	Inclusion stiffness
$\boldsymbol{I}$	Identity tensor
D_i^{dil}	Dilute strain concentration factor
S_i	Eshelby tensor
α, β	Angular orientations
$\bar{\bar{T}}_{klmn}$	Orientationally averaged fourth order tensor
E_{efv}	Effective Young's modulus
E_m	Young's modulus of matrix
κ_{efv}	Effective bulk modulus
μ_{efv}	Effective shear modulus
κ_m	Bulk modulus of matrix
μ_m	Shear modulus of matrix
p, q	Explicit expressions
ρ_{efv}	Effective density
ρ_m	Matrix density
ρ_i	Density of inclusions/reinforcements
F_{xl}	Generalized Newtonian Force in X direction on left-hand side of the reticule
M_{xl}	Generalized Newtonian Moment in X direction on left-hand side of the reticule
F_{xr}	Generalized Newtonian Force in X direction on right-hand side of the reticule
M_{xr}	Generalized Newtonian Moment in X direction on right-hand side of the reticule
x_l	Generalized displacement of left-hand side of the reticule in X direction
x_r	Generalized displacement of right-hand side of the reticule in X direction

φ_l	Generalized angular displacement of left-hand side of the reticule in XZ plane
φ_r	Generalized angular displacement of right-hand side of the reticule in XZ plane
F_{yl}	Generalized Newtonian Force in Y direction on the left-hand side of the reticule
M_{yl}	Generalized Newtonian Moment in Y direction on left-hand side of the reticule
F_{yr}	Generalized Newtonian Force in Y direction on right-hand side of the reticule
M_{yr}	Generalized Newtonian Moment in Y direction on right-hand side of the reticule
y_l	Generalized displacement of left-hand side of the reticule in Y direction
y_r	Generalized displacement of right-hand side of the reticule in Y direction
θ_l	Generalized angular displacement of the left-hand side of the reticule in YZ plane
θ_r	Generalized angular displacement of right-hand side of the reticule in YZ plane
$[K]$	Stiffness matrix
E	Young's modulus
I	Moment of inertia
a	Effective length of the shaft
μ_i	Internal damping coefficient of the shaft
$\dot{X}_r$	Rotary coordinates in X direction
$\dot{Y}_r$	Rotary coordinates in Y direction
∂	Transformation angle
e	Effort through bond
f	Flow through bond
M	Nodal mass
R	Radius of shaft
n	Mode number
R_i	Internal damping
R_{ex}	External damping
C	Stiffness
E_o	Effort output
F_i	Force Input
D_f	Flow detectors
ω_{zz}	Modulus of gyrator element
ρ	Density
Ω	Excitation frequency
I_{phub}	Polar moment of inertia
SF	Source of flow
GY	Gyrator element
R_c	Dissipative coupling coefficient

μ_a	External damping coefficient
ω_n	Natural frequency
ξ	Damping factor
f_1, f_2	Frequencies corresponding to the square root of maximum amplitude value
$f_{\max}$	Frequency corresponding to amplitude value
$\dot{\theta}_t$	Shaft entrainment speed
γ_n	Normalized shaft spin
A_n	nth mode amplitude

Introduction

With the advent of composite materials, industries have started focusing on the use of lighter weight materials with same or higher mechanical properties. Metal matrix composites (MMCs) have an edge over the parent metals for rotor applications as they have a higher specific modulus as specific modulus is the key material factor for vibration responses. Aluminum/alumina MMCs have shown prominent growth in the composite material market because of their compatibility to the rotor systems (Chawla and Shen 2001; Rawal 2001; Sarmasti et al. 2011). Many manufacturers, such as Duralcan, GKN, Toyota, GM, and Nissan, have used MMC materials to develop various components with both static and dynamic applications (Koczak et al. 2013). A 6061Al/Al_2O_3 DRA driveshaft along with cylinder driveshafts manufactured by a stir casting method followed by direct seamless extrusion has various applications (Shercliff and Ashby 1994; Hunt and Miracle 2001; Miracle 2005).

Tailored properties can be achieved by adding reinforcement to a metal matrix in order to form metal matrix composites. Major mechanical properties can be predicted through various methodologies if factors such as composition, aspect ratio, and size of particles are known. Micro-structural images obtained through scanning electron microscopy can be analyzed in order to get the value of these factors.

Multiscale computation of composites is a two-phase process where the effective macroscopic properties are derived through micro-mechanics. Analysis of composites with discrete reinforcements with random orientation has been found in archival literature (Chaturvedi and Tzeng 1991; Guden and Hall 1998; Lesieutre 1993; Vaidya et al. 1997). There are various approaches through which these effective properties can be predicted efficiently (Mura 1987). For composites with a low volume fraction of inclusions, Eshelby's method (Eshelby 1957) with a dilute scheme can be used to evaluate the effective properties. Although for composites with a higher volume fraction of inclusions, aggregate model (Christensen and Malls 1972; Halpin et al. 1971) can be used, which utilizes the known properties of constituents. Mori and Tanaka (1973) have proposed a theory by assuming that the average strain in the inclusion is related to the average strain in the matrix by a fourth-order tensor. The relation between the uniform strains in the inclusion in the matrix is provided by this fourth-order tensor. It was further extended by Benveniste

(1987) to investigate the stress and strain concentration tensors along with the overall elastic modulus of composites. This theory has been adopted by various researches (Abaimov et al. 2019; Imani et al. 2018; Kamarian et al. 2018; Tran et al. 2018; Sobhaniaragh et al. 2017) to determine the effective properties, such as modulus, density, and thermal expansion, and is used in this work.

For various rotor systems, natural frequency, steady-state response, stability, and the transient response passing through the critical speeds are the pivot of dynamic analysis of the rotors. In addition to that, a system approach, which can encapsulate different domains of the energy system, is a prerequisite for the analysis of rotor systems. For that reason, a modular and robust rotordynamic model is required for the study of the response of the rotor system for different internal properties. A bond graph is a multi-domain approach, which is capable to support such an objective. One of its advantages is that it enables the user to visualize the power flow and track the effects of these variants, which can be very useful for understanding the counterintuitive phenomenon of rotor dynamics for any system (Loewy and Piarulli 1969). This approach, coined by Beaman and Paynter (1993), has been adopted in various researches for modeling the different variants of rotor dynamic systems (Margolis 1980; Karnopp 1981). This technique has been used by Hubbard (1979), who has used the bond graph technique and interpreted whirling modes and critical frequencies associated with it. However, no such studies have been attempted to model the composite rotor shaft with internal damping through bond graph modeling. Since the bond graph facilitates the system modeling from the physical paradigm itself, Rastogi et al. (2011) successfully introduced both external and internal damping of the shafts in the bond graph. A detailed description of this technique is provided in Appendix A.

Various archival literature has used this approach to analyze the rotor dynamics of shafts. Dynamics of the rotor systems depend decidedly on the internal damping as well as external damping of the system. External damping as stated is the external friction acted because of the resistance of the medium (liquid/air). In contrary to that, internal damping is the friction existing within the rotating part and is the intrinsic property of the material. Since internal damping is the material property, the dynamics of the rotor changes drastically with the change of material and calibrated internal damping is mandatory to correctly imitate the dynamic response computationally. The internal damping characteristics of conventional metals, predominantly steel in many instances, have been well-documented and utilized in various recent studies focused on the dynamic analysis of shafts (Karthikeyan et al. 2015; Bharti et al. 2019; Jain et al. 2018). However, for composites, internal damping changes with various factors like reinforcement size, aspect ratio, volume percentage, etc. Efforts for determining the effect of these factors on the modal response by utilizing the finite element method are shown in various researches (Chang et al. 2004a, b; Banerjee and Su 2006; Sino et al. 2008). However, such studies are limited to the linear and small variation and have only highlighted the effect of damping on the rotor dynamics of composite shafts. Recently, Kafi and Hosseini (2019) have shown that damping has the main effect on the dynamic response of the composite shafts and analyzed the nonlinear dynamic response of the composite shaft but the

experimental modal characterization is not available. Experimental characterization of metal matrix composites is limited to the evaluation of tensile strength, impact strength, and tribological properties. There are a few types of research available for the dynamic response of composite plates (Souza et al. 2019; Oliveira et al. 2016). While experimental analyses on the modal response of shafts have been limited, recent studies have computationally and experimentally assessed critical speeds for aluminum–boron carbide metal matrix composite shafts (Bajpai et al. 2018). However, they have directly used the effective properties of the MMC shaft and not compared the analytical result. Recently, Allien et al. (2019) have used beams with smaller length specimens for optimization and dynamic analysis of silicon carbide reinforced different aluminum matrix MMCs. However, such smaller length specimens were restricted for rotor drive applications.

To the best of author's knowledge, there is no such literature available recently, which provides experimentally driven internal damping coefficient values, which can be employed in the bond graph modeling. This research work has adopted a novel approach of providing a dedicated methodology for the determination of the calibrated internal damping factor for dynamic analysis through the bond graph, which has been used in various literature for transient and stable responses. The analysis has been performed on long and slender shafts. The next section will highlight the fabrication and experimental response for composite shafts.

Fabrication and Experimental Impact Hammer Response of Different MMC Shafts

Long and slender 6061 Al/Al_2O_3 shafts with different compositions viz. 4%, 8%, and 12% of alumina (Al_2O_3) by volume, with a dimension of 22 mm $\times$ 1000 mm were fabricated through stir casting. Optimized parameters for low porosity and better homogeneity have been extracted for different size and composition of the product by many types of research (Ravi et al. 2007; Naher et al. 2003; Singh et al. 2017; Sahu and Sahu 2017; Assar et al. 1982; Hashim et al. 2002). Sahu and Sahu (2018) provided generic optimized data influenced by these researches, which has been used in this work. Vortex was generated through stirring, by a four-blade stirrer with 30° angle, rotating at 550 rpm up to 10 min.

Vibration analysis of these sample shafts was performed by an experimental rig, as shown in Fig. 1. A four-pole, single-phase DC motor was employed to drive the shaft. It was coupled to the shaft through a flexible and dissipative coupling with flexible neoprene spider mounted between the couplings. Shafts were supported by self-aligned 1204 K C3 NTN bearings with 22 mm internal diameter at a distance of 660 mm. Test specimens were of 22 mm diameter, effective length (Bearing to bearing) was 660 mm, and the shaft was overhung by 175 mm from each end. The vibrational responses (static and dynamic) of the test specimen were recorded and analyzed by using a multi-port vibration analyzer OROS36®. The channels of these

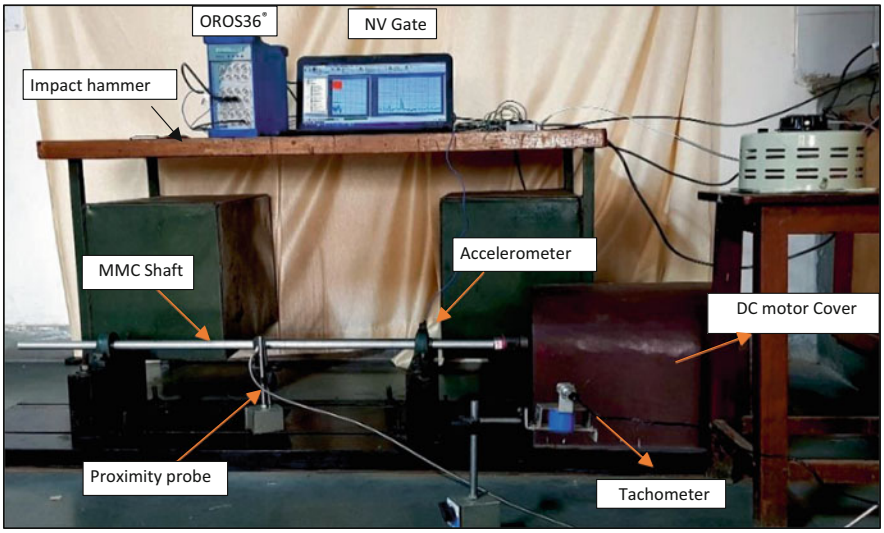


Fig. 1 Experimental set up for dynamic and impact hammer analysis

analyzers are handled in real-time: FFT, 1/3rd Octave. NV Gate® software was used as an interface between the system and analyzer. NV Gate platform offers a comprehensive set of tools for noise and vibration acquisition, recording and analysis.

For impact hammer analysis, vibrational responses were captured through an inductive proximity sensor (LJ12A3-4-Z/BY) with a detection range of 4 mm and sensitivity of 1 m(V)/m(m). This sensor was mounted on a support and was oriented in such a manner that the tip of the sensor was at a distance of 2 mm from the diametrical end of the shaft. Impact Hammer (Model: – PCB-086C03) with sensitivity 2.25 m (V)/(N) and range 4.44kN was used for transient excitation. Also, the rubber tip was used for the impact since test specimens were of a softer material than that of steel (metal tip material). In this experiment, sample rate taken is 2.04 kHz/s. In general, the Hanning (Hann) window is satisfactory in 95% of cases. It has good frequency resolution and reduced spectral leakage. Therefore, a Hanning window was used to perform the time windowing of the signals in both static and dynamic analyses. Since impact load itself results in minimal leakage, therefore, time weighing does not impact on the results.

Several iterations were taken so that the initial displacement of 300 microns for the analysis was equal in all cases and the FFT was captured after 10 s of the impact. The response of the different shafts rotor dynamics below.

The details of MMC shafts impact hammer response along are in Fig. 2 with the damping ratio calculated by half-power bandwidth are shown in Table 1:

The experimental impact hammer response depicts that the natural frequency increases and the amplitude decreases with the increase of the reinforcement percentage. This trend is congruent with the archival literature on experimental responses (Bajpai et al. 2018; Allien et al. 2019).

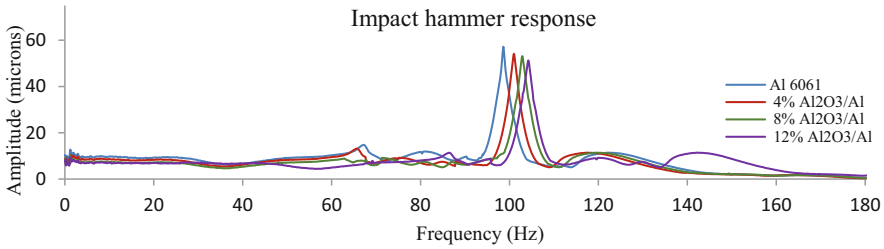


Fig. 2 The response of impact hammer analysis

Table 1 Values of amplitudes, damping ratio, and frequencies from impact hammer response

Material	Amplitude of vibrations	Damping ratio	Frequency
Al 6061	58.3	9.511×10^{-3}	98 Hz
4% Al_2O_3/Al 6061	55.2	9.812×10^{-3}	101.125 Hz
8% Al_2O_3/Al 6061	54.1	1.007×10^{-3}	102 Hz
12% Al_2O_3/Al 6061	52.4	1.112×10^{-3}	104.25 Hz

Estimation of Effective Mechanical Properties Through Mean-Field Homogenization by Using Parameters Extracted Through Image Processing

Cubic samples of 5 mm length, 5 mm width, and 5 mm height were processed by EDM wire cutting, and scanning electron microscopy (SEM) micrograph images at the magnification of 500 and 1000 were synthesized by a Zeiss EVO 50 machine. The morphology of the micrograph clearly depicts the uniform distribution of reinforcements in the aluminum matrix. Agglomeration was induced for a higher percentage of the reinforcement but to a limited extent, where porosity may not play a vital role in determining mechanical properties of the material except for damping characteristics. For each composition, three samples were scanned and fraction/size of particles was evaluated by image processing and averaged for the computation of effective properties.

The open-source image processing software ImageJ was used to create binary images of the composites in the ellipsoid form (since for homogenization of property ellipsoidal particles are assumed). Size of the reinforcement, aspect ratio, and volume percentage for the composite were derived from the image analysis, as shown in Fig. 3. These parameters were incorporated in the mean-field homogenization of the metal matrix composite material where a representative volume element (RVE) is assumed to be the building block of the macroscopic structure. The effective properties of these RVE's were calculated and have been used in the bond graph for further dynamic analysis.

For mean-field homogenization, Mori-Tanaka method (Mori and Tanaka 1973) with Wang and Tandon model (Mukherjee and Samantaray 2000) has been adopted

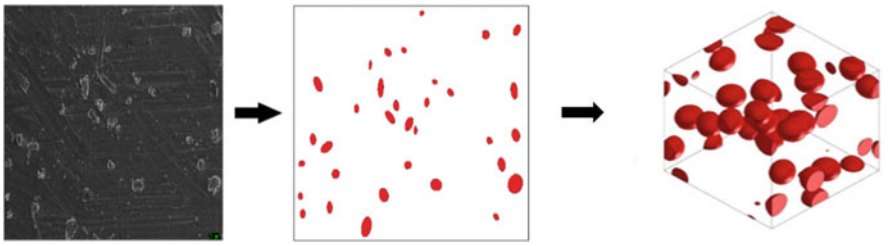


Fig. 3 Image processing and conversion into RVE

to determine the effective Young's modulus of the composite material. Therefore, it assumed that the metal matrix composite is comprised of N number of phases. Matrix phase is denoted by subscript m, and there are remaining N-1 inclusion phases in the control volume. The volume fractions of the matrix and the ith inclusion are denoted by ϑ_m and ϑ_i. Similarly, stiffnesses of the matrix and the ith inclusion are denoted by K_m and K_i. These stiffness quantities are generally represented as elasticity tensors of fourth-order, which possess certain symmetrical properties, i.e., $K_{klmn} = K_{lkmn} = K_{klnm}$ and $K_{klmn} = K_{mnkl}$.

The overall elastic stiffness of the composite (Benveniste 1987) is defined as:

$$\bar{\bar{K}}_{av} = \left(\vartheta_m * K_m + \sum_{i=1}^{N-1} \vartheta_i \left\langle K_i * D_i^{dil} \right\rangle\right) + \left(\vartheta_m * \boldsymbol{I} + \sum_{i=1}^{N-1} \vartheta_i \left\langle D_i^{dil} \right\rangle\right)^{-1} \tag{1}$$

where $\boldsymbol{I}$ is the identity tensor and the dilute strain concentration factor D_i^{dil} is generally expressed as:

$$D_i^{dil} = \left[\boldsymbol{I} + S_i K_m^{-1}(K_i - K_m)\right]^{-1} \tag{2}$$

Here S_i is the Eshelby tensor (Eshelby 1957) for the spheroidal inclusion in the infinite matrix.

Since the alumina particles are randomly oriented, orientation averaging has to be taken into account. Therefore, the terms in the angle parentheses $\langle\rangle$ denote the averaged value of the parameters over all possible orientations. This orientationally averaged tensor is a fourth-order tensor in the 3D space (Chang et al. 2004a) and can be expressed as:

$$\bar{\bar{T}}_{klmn} = \langle T_{klmn} \rangle = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{2\pi} T_{klmn}(\alpha, \beta) \sin\beta \, d\alpha \, d\beta \tag{3}$$

For the randomization, the local coordinates should be transformed to the global coordinates. For overall random transformation, the matrix may be expressed as:

$$a_{kl} = \begin{bmatrix} \cos\alpha & \cos\beta\sin\alpha & \sin\beta\sin\alpha \\ -\sin\alpha & \cos\beta\cos\alpha & \sin\beta\cos\alpha \\ 0 & -\sin\beta & \cos\beta \end{bmatrix} \tag{4}$$

And the transformation may be written as:

$$T_{klmn}\,(\alpha,\beta) = a_{ko}a_{lp}a_{mq}a_{nr}\,T'_{opqr} \tag{5}$$

The effective Young's modulus (E_{efv}) (Chang et al. 2004a) can be now calculated by the general equation with shear modulus and bulk modulus as:

$$E_{efv} = E_m * \frac{\kappa_{efv} * \mu_{efv}(3\kappa_m + \mu_m)}{3k_{efv} * \kappa_m + \mu_{efv} * \mu_m} \tag{6}$$

Effective Shear modulus is given by:

$$\frac{\kappa_{efv}}{\kappa_m} = \frac{1}{1 + z\,p} \tag{7}$$

Effective Bulk modulus is given by:

$$\frac{\mu_{efv}}{\mu_m} = \frac{1}{1 + z\,q} \tag{8}$$

where κ_{efv} and μ_{efv} are the effective bulk and shear moduli of the composite, z is the volume concentration/fraction, κ_m and μ_m are the bulk and shear modulus of the matrix. Also, p and q are the explicit expressions, and these expressions and details of the expression can be seen in Tandon and Weng (1986) for spheroidal inclusions for detailed reference.

Also, effective density can be easily calculated through the rule of the mixture as:

$$\rho_{efv} = \rho_m\vartheta_m + \rho_i\vartheta_i \tag{9}$$

where ρ_{efv} is the effective density, ρ_m and ϑ_m are the density and volume fraction of matrix, and ρ_i and ϑ_i are the density and volume fraction of inclusions.

The collective results for different combinations of metal matrix composite shafts are provided in Table 2:

The next section will highlight the modeling of the composite shaft through bond graph modeling.

Bond Graph Modeling

The physical system of a simply supported MMC shaft coupled to a DC motor with a dissipative coupling is shown in Fig. 4. The shaft's material is assumed to be homogeneous and continuous throughout the length. The beam is assumed to be simply supported at the ends.

Table 2 Homogenized effective properties

Material	Average volume fraction	Average aspect ratio	Average size of alumina (microns)	E_{efv} (GPa)	ρ_{efv} (Kg/m^3)
Al 6061	–	–	–	70	2700
4% Al_2O_3/ Al 6061	4.101	2.8	8.1	73.73	2748
8% Al_2O_3/ Al 6061	8.098	3.4	8.3	77.66	2796
12% Al_2O_3/ Al 6061	12.23	3.2	9.2	81.785	2844

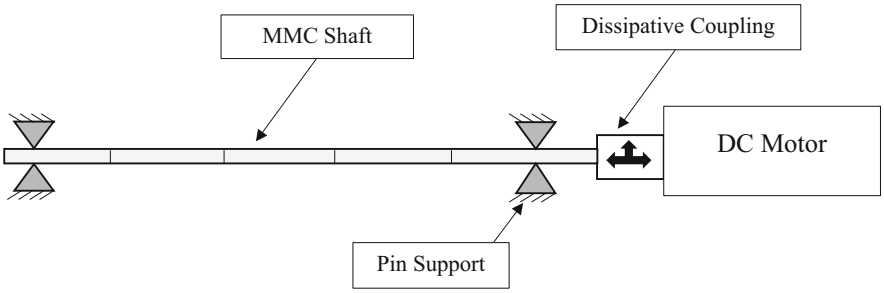


Fig. 4 Rotor system with DC motor drive and pin-pin end condition

The effect of torsional vibrations on the shaft response is neglected in this modeling. The shaft of the composite is assumed to be a Rayleigh beam model where beam's rotary inertia is considered in the analysis. The Newtonian sign conventions are assumed in this work, as shown in Fig. 5. Moreover, it is assumed that the segments of the reticulated beam store the potential energy because of displacements at the ends. There are four displacements at the ends where two of them are linear displacements and other are rotational displacements.

The generalized Newtonian forces are related to the generalized displacements at the ends of the element as:

$$\begin{bmatrix} F_{xl} \\ M_{xl} \\ F_{xr} \\ M_{xr} \end{bmatrix} = [K] \begin{bmatrix} x_l \\ \varphi_l \\ x_r \\ \varphi_r \end{bmatrix} \tag{10}$$

F_{xl} is the shear force due to the displacement at x_l, i.e., displacement of the left end in the x-direction. M_{xl} is the bending moment due to the rotation φ_l. Subscript l is for the left end and r is for the right end of the element.

For modeling through the bond graph, a four-port C- field junction represents the potential energy storage as shown in Fig. 6. The corresponding rotational and linear

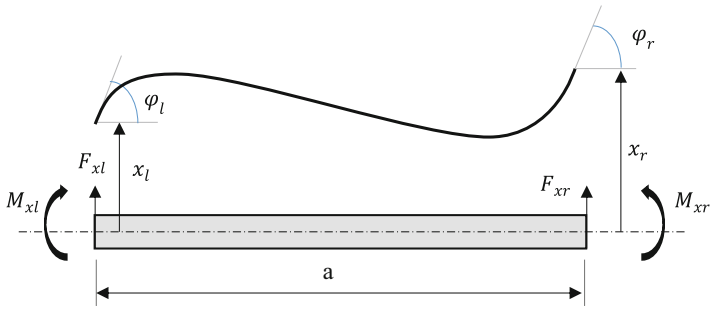


Fig. 5 Force and moment on a Rayleigh beam in XZ plane

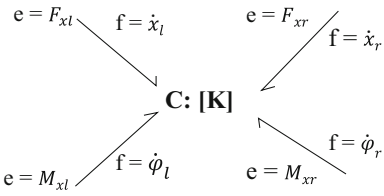


Fig. 6 Representation of bond graph of element stiffness

velocities are the flow variables for the ports of the storage C- fields. Similarly, bending moments and shear forces are the effort variables, as shown in Fig. 6.

The stiffness matrix may be derived by various methods and obtained as:

$$[K] = \frac{EI}{a^3}\begin{bmatrix} 12 & 6a & -12 & 6a \\ 6a & 4a^2 & -6a & 2a^2 \\ -12 & -6a & 12 & -6a \\ 6a & 2a^2 & -6a & 4a^2 \end{bmatrix} \tag{11}$$

The stiffness matrix is a symmetric matrix where EI is the flexural rigidity and a is the effective length of the shaft.

Similarly, the relations can be derived for XZ and YZ direction as:

$$\begin{bmatrix} F_{xl} \\ M_{xl} \\ F_{xr} \\ M_{xr} \end{bmatrix} = \frac{EI}{a^3}\begin{bmatrix} 12 & 6a & -12 & 6a \\ 6a & 4a^2 & -6a & 2a^2 \\ -12 & -6a & 12 & -6a \\ 6a & 2a^2 & -6a & 4a^2 \end{bmatrix} \begin{bmatrix} x_l \\ \varphi_l \\ x_r \\ \varphi_r \end{bmatrix} \tag{12}$$

$$\begin{bmatrix} F_{yl} \\ M_{yl} \\ F_{yr} \\ M_{yr} \end{bmatrix} = \frac{EI}{a^3}\begin{bmatrix} 12 & 6a & -12 & 6a \\ 6a & 4a^2 & -6a & 2a^2 \\ -12 & -6a & 12 & -6a \\ 6a & 2a^2 & -6a & 4a^2 \end{bmatrix} \begin{bmatrix} y_l \\ \theta_l \\ y_r \\ \theta_r \end{bmatrix} \tag{13}$$

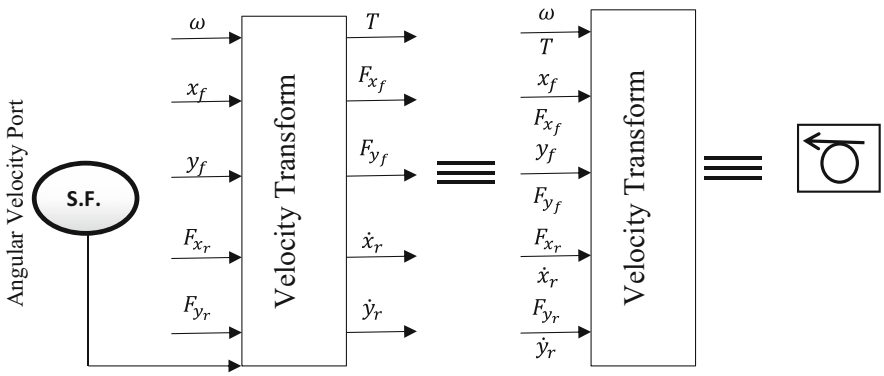


Fig. 7 Velocity transformation capsule

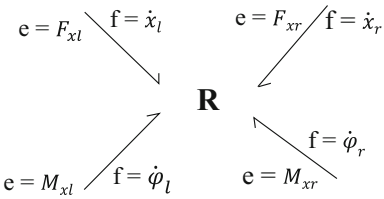


Fig. 8 Representation of element internal damping

For bond graph modeling, the shaft is reticulated in ten equal parts and the element inertias are lumped at the left and right of each reticulated part. For modeling the stiffness matrix, four-port compliance fields (C fields) have been denoted as an energy-storing element for the different generalized displacements. Since the internal damping force is related to the inertial velocity of the mass center by observing in the rotary frame, the rotary generalized displacements should be linked to the internal damping resistance fields (R fields). For that, a transformation sub-model has been generated to transform the generalized displacements from fixed to rotary frame velocity in the form of a capsule. The capsule has been generated by taking the following relation into consideration.

$$\begin{bmatrix} \dot{X}_r \\ \dot{Y}_r \end{bmatrix} = \begin{bmatrix} \cos\partial & \sin\partial \\ -\sin\partial & \cos\partial \end{bmatrix} \begin{bmatrix} \dot{X}_f \\ \dot{Y}_f \end{bmatrix} - \omega \begin{bmatrix} \sin\partial & -\cos\partial \\ \cos\partial & \sin\partial \end{bmatrix} \begin{Bmatrix} X_f \\ Y_f \end{Bmatrix} \tag{14}$$

and $\dot{\partial} = \omega$

The Velocity transformation capsule is represented in Fig. 7 as:

The transformed rotary generalized displacements rates are connected to a four-port resistance field (R field) for appending the internal damping effect into the dynamics of the MMC shaft, as shown in Fig. 8.

The corresponding damping matrix may be expressed as:

$$[R] = \frac{\mu_i EI}{a^3}\begin{bmatrix} 12 & 6a & -12 & 6a \\ 6a & 4a^2 & -6a & 2a^2 \\ -12 & -6a & 12 & -6a \\ 6a & 2a^2 & -6a & 4a^2 \end{bmatrix} \tag{15}$$

where EI is the flexural rigidity, a is the effective length of the shaft, and μ_i is the internal damping coefficient which is the intrinsic property of the material. This value of intrinsic damping coefficient has been calibrated by mapping the bond graph modeling response by the experimental response.

The spinning shaft capsule for a reticule has been generated, as shown in Fig. 9. The external damping (R_{ex}) is associated with the fixed X and Y frame. Common flow junctions (1 junction) for generalized displacements in fixed and rotary frames are linked to the corresponding R fields and C fields. A finite number of reticules can be applied to different applications.

Angular velocity port is introduced in the velocity transform capsule for importing the angular speed into the model. Flow input interface (F_i) is connected to the left side of the fixed generalized velocity junctions ($\dot{x}_l, \dot{y}_l, \dot{\theta}_l, \dot{\varphi}_l$). Similarly, the effort output interface (E_o) is connected to the right side of the fixed generalized velocity junctions ($\dot{x}_r, \dot{y}_r, \dot{\theta}_r, \dot{\varphi}_r$). These interfaces are also connected to the corresponding velocity transform where the fixed generalized velocities are converted to rotary frame velocities and incorporated in the R fields. As there is no proportionality between the rotations between the angle of rotation about X-axis and the ends of the shaft due to bending in YZ-planes as used in deriving C-field matrix, two-transformer elements (TF) with negative unitary modulus are introduced in the model, as shown in Fig. 9.

This spinning shaft model is canned into a capsule so that it can be used for multiple port analysis, as shown in Fig. 10.

The lumped masses and rotary inertias at the ends of the reticulated shafts have to be connected by a spinning hub for connection of flow input port to the effort output port of the element. The spinning hub models the inertia of the shaft element. If required, it may also incorporate eccentricity and the angular orientation of the hub center of mass. The angular speed of the spinning hub is also induced separately in the model, and the spinning hub model is converted to a capsule, as shown in Fig. 11. The D_f elements in the model are flow detectors, which provide the output variables for plotting whirl orbits, whirl amplitudes, modes shapes, etc.

The nodal mass at any intermediate node for a model with n elements is $M = \frac{\pi R^2 a}{n} * \rho$, a diametric moment of inertia $I_{xx}, I_{yy} = \frac{\pi R^4 l}{4} * \rho$. The translational damping (R_t) indicates the resistive elements in fixed X- and Y -frames, whereas rotational damping (R_{rot}) represents about the diametrical axis resistive elements in rotating X- and Y-frames. The modulus of the gyrator element in the hub model may be expressed as:

$$\omega_{zz} = \left\{\frac{\rho\pi}{2}R^4 + I_{\text{phub}}\right\}\Omega \tag{16}$$

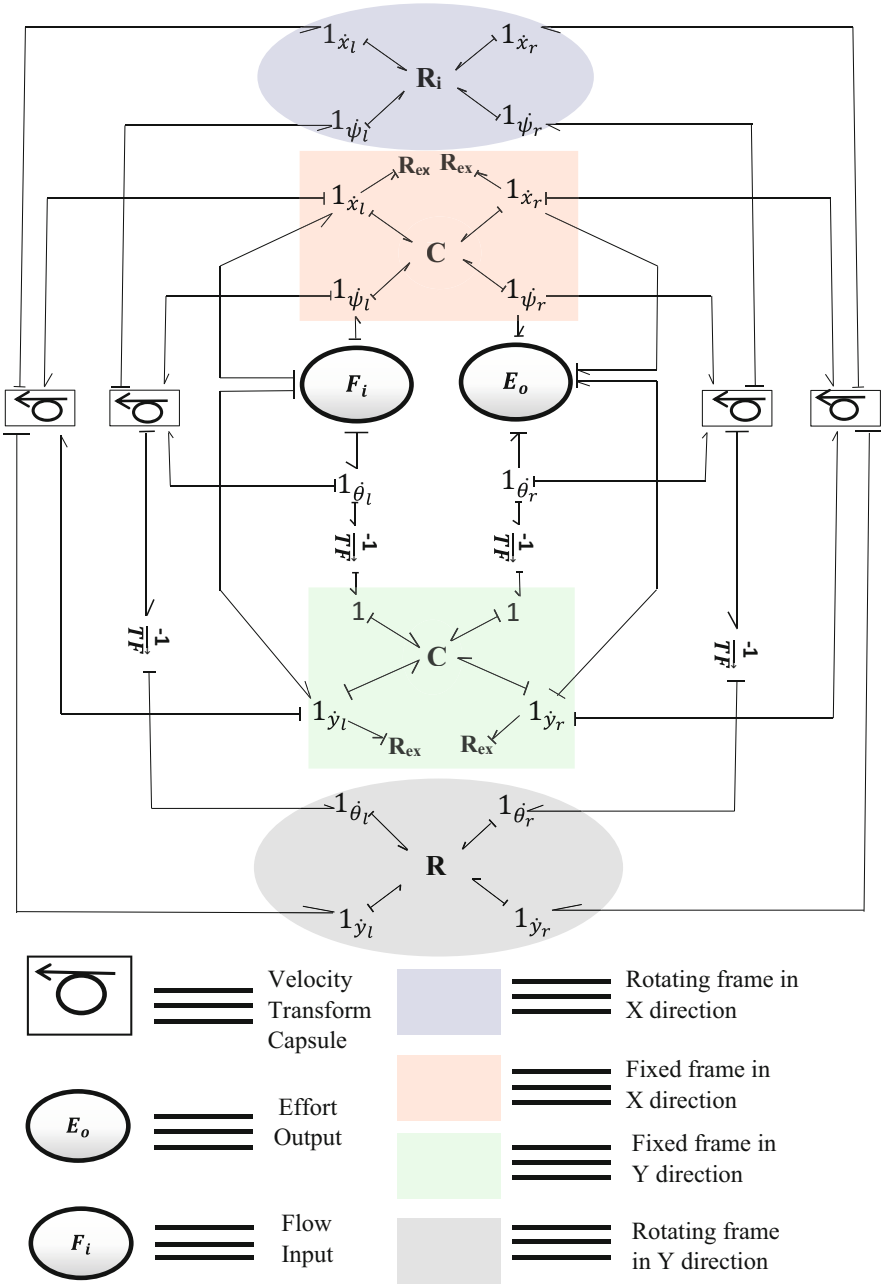


Fig. 9 Insight of the composite spinning shaft mechanism

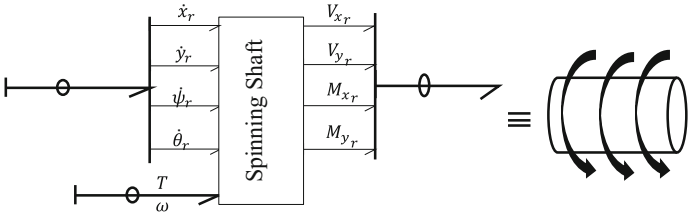


Fig. 10 Spinning shaft capsule

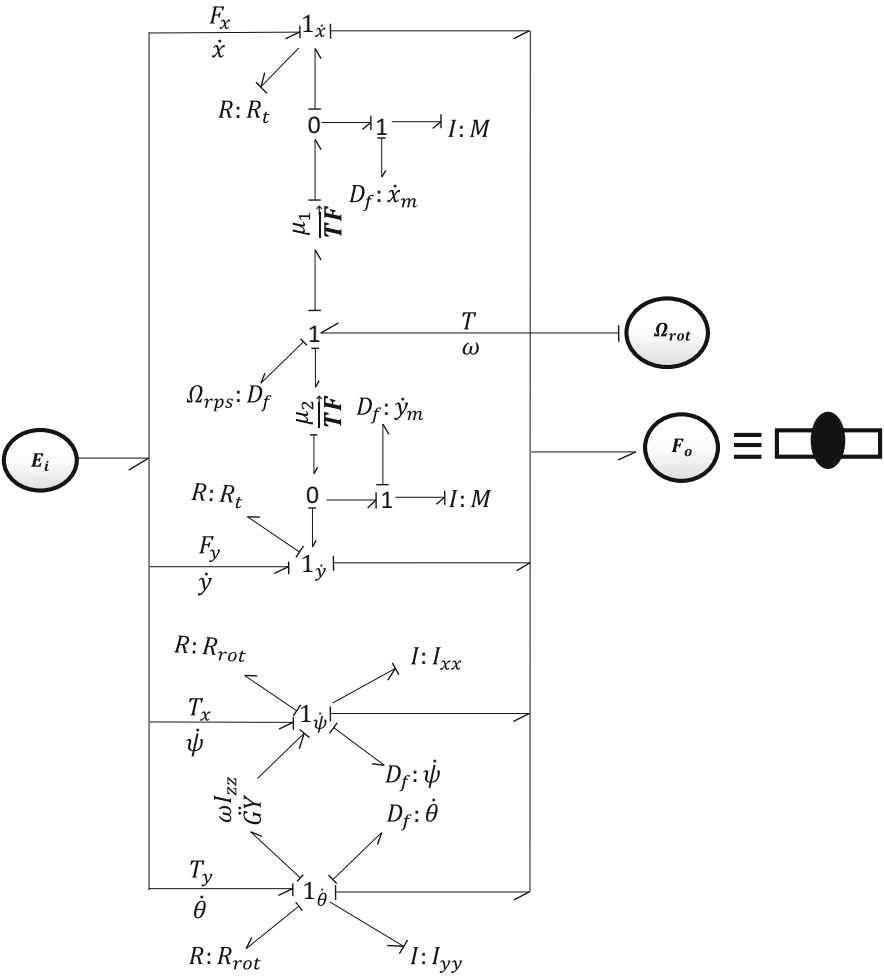


Fig. 11 Spinning hub capsule

where I_{phub} is the polar moment of inertia of the hub and Ω is excitation frequency of the shaft. The transformer function elements have their moduli, as shown in Fig. 11, and their values are:

$$\mu_1 = -\varepsilon \sin(\Omega t + \emptyset) \tag{17}$$

$$\mu_2 = \varepsilon \cos(\Omega t + \emptyset) \tag{18}$$

where ε is the eccentricity of the hub and $\emptyset$ is the angle of equivalent single unbalanced mass measured from the vertical axis.

To integrate and interface the different sub-models of the various parts of the rotor systems, the vectorized bond graph has been employed, as shown in Fig. 12. The shaft is reticulated in ten different elements with nine spinning hub elements between them. The dissipative coupling between the shaft and the drive motor is denoted as R_c, and the angular rotation of the shaft is expressed as Ω. The total polar moment of inertia of the shaft can be represented as:

$$I_p = \frac{\pi\rho}{8} R^2 * \text{Total length of shaft} \tag{19}$$

the end condition is assumed to be pin-pin, where the source of flow is zero and the inertial element is:

$$I_d = \frac{\pi\rho}{8} L_{\text{elem}} R^4 \tag{20}$$

where L_{elem} is the length of the reticulate element.

In the present model, the value of the polar moment of inertia, a moment of inertia, and mass of hub is considered as zero. The pin-pin end supports are assumed to be isotropic and rigid.

Simulation

The simulation was performed on a dedicated bond graph software Symbols Shakti Software (Hashim et al. 2002) using a fourth-order Runge-Kutta method. All the object-oriented sub-models are connected on bondpad, as shown in Fig. 12.

The parameters such as dissipative coupling coefficient, external damping coefficient of the rotor shaft (μ_a), and number of elements are driven from the literature (Rastogi et al. 2011) and the dimensions of the shaft are as per the experimental setup whose values are shown in Table 3.

Parameters that change with the material are Young's modulus, density, and internal damping. The values of Young's modulus and density for different materials were calculated through mean-field homogenization with values as provided in Table 2.

As reported in various observations and literature, internal damping effects the dynamic behavior of the rotating shafts. A phenomenon where the presence of

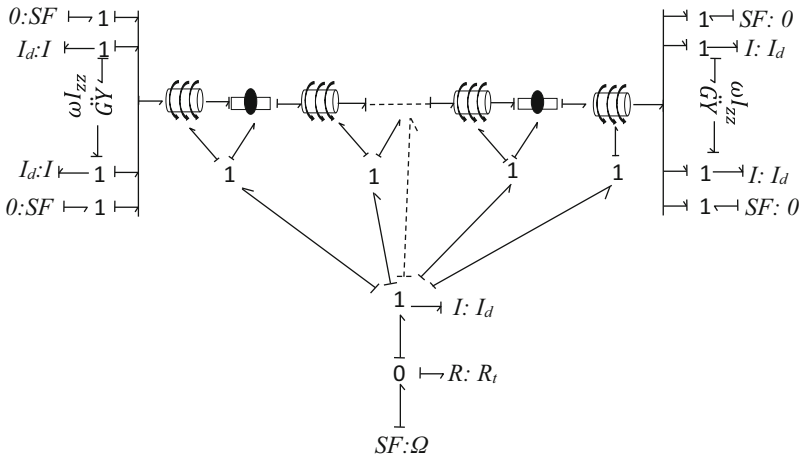


Fig. 12 Bond graph model by connecting sub-models

Table 3 Parameters for simulation

Parameter	Value
Effective length of the shaft	$L_{\text{beam}} = 660$ mm
Radius of the shaft	$R_o = 11$ mm
Number of reticulated elements	$N_{\text{elem}} = 10$
External damping coefficient of the shaft[a]	$\mu_a = 3.00\text{e}^{-5}$, 6.00e^{-4} Ns/m (for modes 1, 2 respectively)
Dissipative coupling coefficient	$R_c = 0.002$ Nms

[a]Damping coefficients are multiplied with EI besides these value

external and internal damping results in the instability of the shaft above a certain speed depends on the ratio of the external and internal damping. Also, changing the material of the shaft does not affect the values of external damping significantly. Additionally, the alteration in material brings about an effective change in internal damping. So, for simulation calibration of the internal damping coefficient of the shaft has been performed in the present work.

Now, the approximation for the internal damping coefficient (μ_i) has been taken as:

Approximate value of internal damping coefficient of shaft* $= 2m\omega_n\, x\, \xi = 1.3 \times 10^{-3}$ Ns/m.

*Damping coefficients are multiplied with EI besides these values.

This approximation is based on an assumption of considering the system having a single degree of freedom and the source and vectorial position of the damping cannot be identified. So, it is customary to include damping functionally into the bond graph (Margolis 1980). To incorporate this, the calibrated internal damping is derived by connecting it to the natural response of the rotor.

For calibration of the internal damping coefficient for simulation through the bond graph, the damping ratio (ξ) calculated experimentally has been mapped with the damping ratio obtained by the bond graph response. Half-power bandwidth formulation has been used for both the calculations. According to half-power bandwidth formulation, the damping ratio corresponding to these values may be evaluated as

$$\text{Damping factor/ratio}(\xi) = \frac{f_1 - f_2}{2 \times f_{\max}} \tag{21}$$

where $f_{\max}$ is frequency corresponding to the maximum amplitude value of FRF plots and f_1 and f_2 are the frequencies where the values are the square root of the maximum amplitude.

For bond graph analysis, the Fast Fourier Transform (FFT) response was extracted by providing an initial displacement of 300 microns and 10 s delay (same as experimental analysis). As shown in Fig. 13 for Al 6061, there was a huge difference between the damping factors of the experimental and bond graph data. For the other MMC shafts, the same damping coefficient was taken, and the results are shown in Table 4.

As shown in Table 4, the percentage error reduced for composites with a higher volume percentage of reinforcement. This clearly depicts that the approximated coefficient of internal damping generated the results with higher damping ratio (which indicates the internal material damping characteristics). So, the internal damping coefficient (μ_i) should be reduced in order to get the proper result. Moreover, the internal damping coefficient (μ_i) for every MMC compositions will change accordingly. The calibrated values of the internal damping coefficient (μ_i) are shown in Table 5, and the mapped graph plot for all compositions is shown in Fig. 14.

Now the overall response of the static and rotatory response of MMCs shaft can be extracted through bond graph. Example of Al 6061 shaft displacement response with the calibrated internal damping coefficient (μ_i) is shown in Fig. 15.

Figure 15 shows the real part, imaginary part, amplitude value and phase of the FFT response generated by the bond graph. The real part is even symmetric, and the imaginary part is approximately odd symmetric. The phase change occurs at the frequency of 100 Hz and 400 Hz, which are in proximity to the experimental response. The amplitude conforms the resonating effect occurred in the experimental results. Since the values are derived from the first mode of response, the focus is on the first mode of resonance. The bond graph response with initial displacement is shown in Fig. 16.

The amplitudes and frequencies of analytical and experimental values are shown in Table 6.

The trend of amplitude and natural frequencies of the results obtained through the bond graph was congruent with the experimental results. The amplitudes were reduced for the MMC shafts with a higher volume fraction of reinforcements due to the increase in Young's modulus and stiffness of the shaft. The natural frequency increased for the higher for the MMC shafts with a higher volume fraction of reinforcements due to the increase in the specific modulus of the MMCs. While

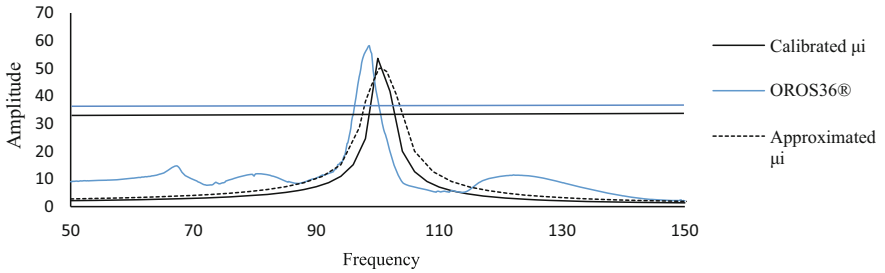


Fig. 13 Internal damping calibration

Table 4 Experimental and bond graph damping ratio for an approximate value

Material	Damping ratio (Experimental)	Damping ratio (Bond graph approximation)	% Error
Al 6061	9.511×10^{-3}	1.12×10^{-3}	15.08%
4% Al_2O_3/Al 6061	9.812×10^{-3}	1.12×10^{-3}	12.31%
8% Al_2O_3/Al 6061	1.007×10^{-3}	1.12×10^{-3}	10.08%
12% Al_2O_3/Al 6061	1.05×10^{-3}	1.12×10^{-3}	6.25%

there are slight variations between the experimental and bond graph values of amplitude and natural frequencies. The value of the amplitudes of the bond graph analysis is less as compared to the experimental values. In addition to that, the frequency in the bond graph results is higher than the experimental results. The reason for this may be the porosity induced in the MMC due to the agglomeration of particles and stir casting effect.

The results reveal that better dynamic properties can be attained by increasing the reinforcement percentage into the metal matrix. For further analysis of rotating MMC shafts, the bond graph analysis has been extended to the first and second mode of vibrations. The dynamic response of the shafts is characterized by a phenomenon where the spinning speed of the shaft is entrained to a certain limit. Beyond this, the power induced in the system increases the amplitudes of the transverse vibrations (Rastogi et al. 2011; Karthikeyan et al. 2015). Moreover, whirl speed of the shaft is dependent on the natural frequency, and the ratio of the damping coefficients may be obtained as follows:

$$\dot{\theta}_t \leq \Omega_n \left(1 + \frac{\mu_a}{\gamma_n}\right) \tag{22}$$

where $\dot{\theta}_t$ is the shaft entrainment speed, Ω_n is the natural frequency, $\gamma_n = \mu_i \left(\frac{n}{\pi a}\right)^4$, μ_a and μ_i are the coefficient of internal and external damping, n is the mode number, and a is the length of the shaft and corresponding amplitude as

Table 5 Calibrated internal damping factor for different MMC shaft

Material	Calibrated μ_i	Damping ratio (Experimental)
Al 6061	6.1×10^{-4}	9.5×10^{-3}
4% Al_2O_3/Al 6061	6.49×10^{-4}	9.8×10^{-3}
8% Al_2O_3/Al 6061	6.97×10^{-4}	1.0×10^{-3}
12% Al_2O_3/Al 6061	7.6×10^{-4}	1.06×10^{-3}

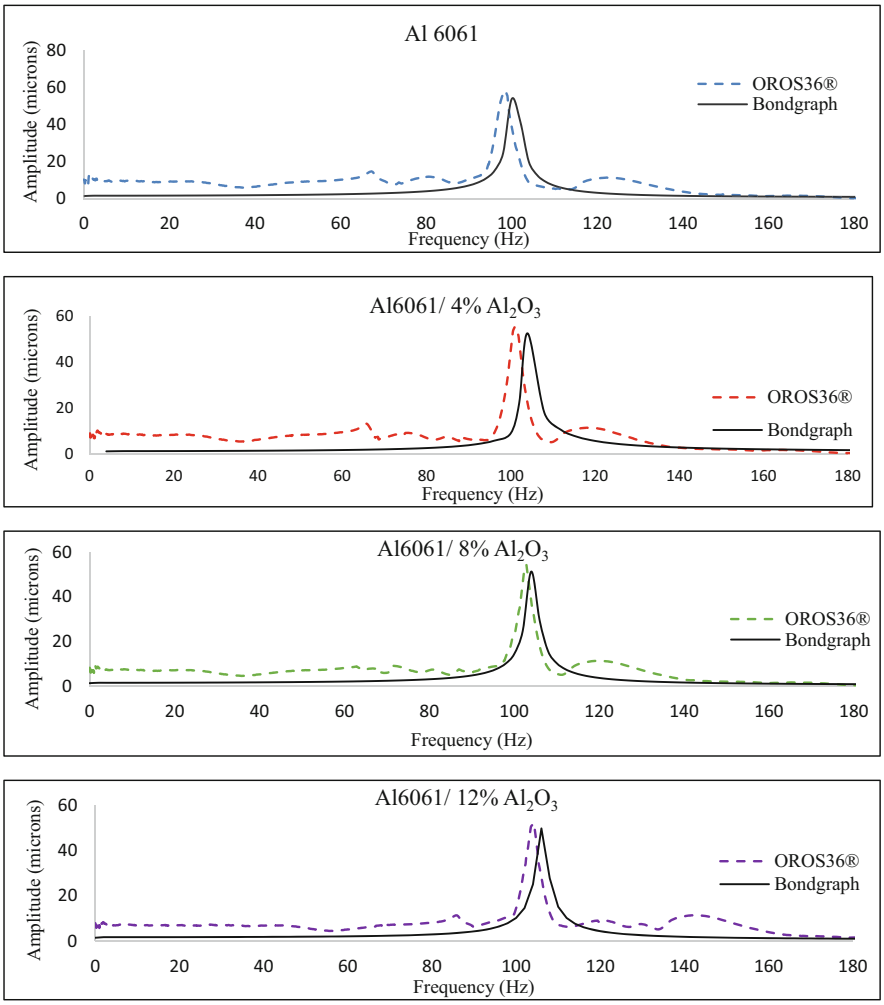


Fig. 14 Mapped plots for calibrated internal damping coefficients

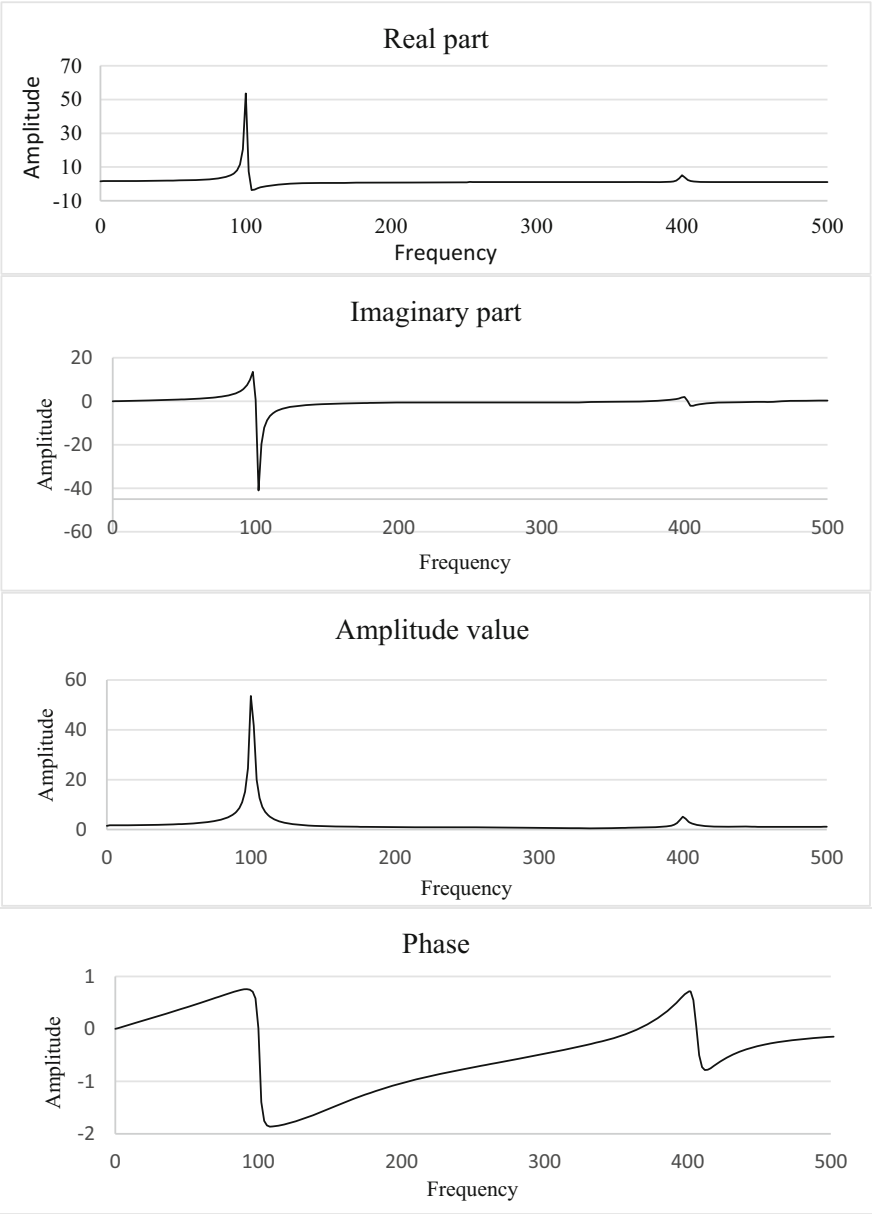


Fig. 15 Displacement response for modeling impact hammer response

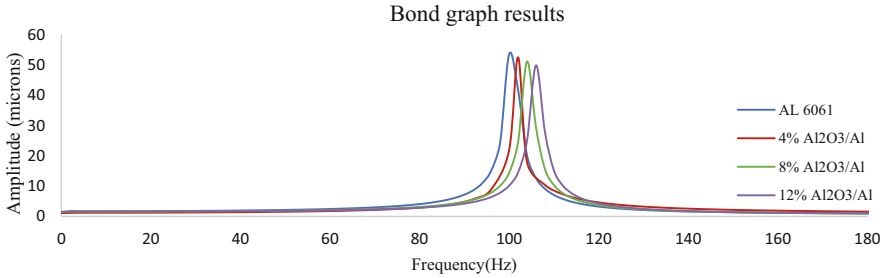


Fig. 16 Response for first mode excitation in bond graph modeling

Table 6 Comparison of the result after calibrated internal damping factor for different MMC shaft

	Amplitude		Natural frequency	
Material	experimental	Bond graph	experimental	Bond graph
Al 6061	58.3	53.6	98 Hz	100 Hz
4% Al_2O_3/Al 6061	55.2	52.501	101.125 Hz	102 Hz
8% Al_2O_3/Al 6061	54.1	51.2	102 Hz	104 Hz
12% Al_2O_3/Al 6061	52.4	49.9	104.25 Hz	106 Hz

$$A_n = \sqrt{\frac{R_c\left(\Omega - \dot{\theta}_t\right)}{\gamma_n \Omega_n}} \tag{23}$$

A_n is the nth mode amplitude and R_c is the dissipative coupling resistance.

For simulation results through bond graph, the excitation speed has been taken as twice the first natural frequency for mode 1, and twice the second natural frequency for mode 2. The shaft entrainment for both the cases was near the natural frequencies congruent to the Eq. (22). The whirl orbit for the response of the MMC shafts in the first and second mode is shown in Fig. 17.

The amplitudes of the whirl orbits reduced for composites with a higher volume fraction of reinforcement. The maximum first mode horizontal and vertical displacements were reduced by 0.44 mm and 0.26 mm, and the maximum second mode horizontal and vertical displacements were reduced by 0.310 mm and 0.18 mm due to the addition of reinforcement. This may be due to the fact that the internal damping coefficient of the shaft increased and the amplitude is inversely proportional to it. Moreover, natural frequency increases with increased reinforcement percentage resulting in a lower difference between the excitation and the entrained frequency. Although entrainment speed is related to the internal damping coefficient and may have affected the amplitude response, the shaft is not long enough that this factor may dictate the shaft entrainment. In the second mode, the amplitudes reduced but the trend remained likely for the same reason.

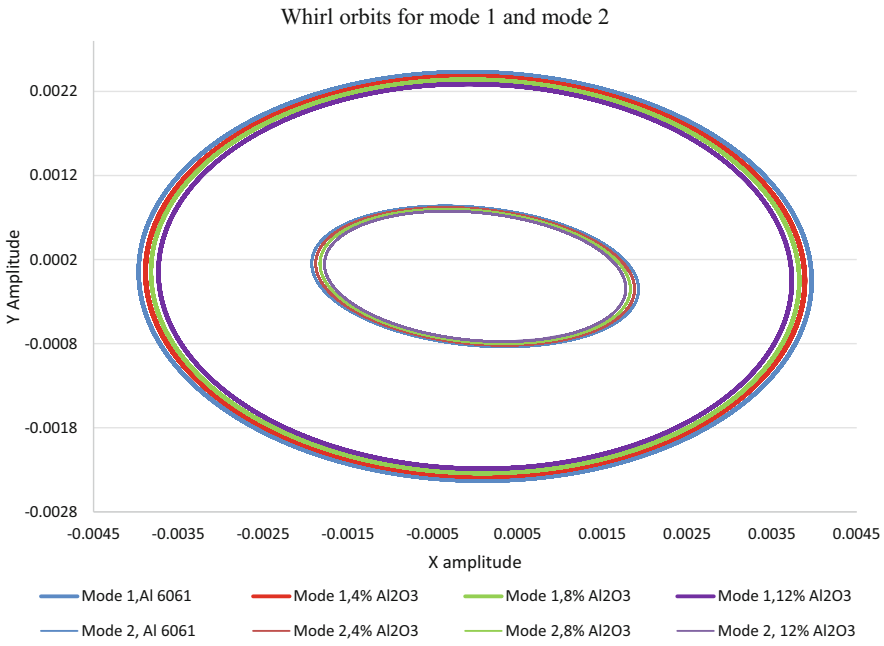


Fig. 17 Whirl orbits or the response of the MMC shafts in the first and second mode

Conclusion

In this chapter, the impact hammer test has been performed to determine the natural frequency, the amplitude of vibrations, and the damping ratio of the stir cast MMC shafts. Reduced amplitudes of vibration were observed for the shafts with a higher volume fraction of reinforcements. Also, it has been noticed that the damping and natural frequency of the rotor systems have been increased with an addition of reinforcements in composites.

Image processing of SEM Images confirmed the homogeneous distribution of alumina particle. The average particle size varies from 2 μm to 23 μm with an average of 8.1 μm for 4% Al_2O_3/Al 6061 composite, 8.3 μm for 8% Al_2O_3/Al 6061 composite, and 9.2 μm for 12% Al_2O_3/Al 6061 composite. Average aspect ratio values were 2.8 for 4% Al_2O_3/Al 6061 composite, 3.4 for 8% Al_2O_3/Al 6061 composite, and 3.2 for 12% Al_2O_3/Al 6061 composite. These parameters were employed to evaluate the effective Young's modulus and density through mean-field homogenization. A multiscale modeling approach was adopted, where these effective properties were evaluated through mean-field homogenization and simulation of the dynamics of composite shafts through bond graph modeling.

Bond graph simulations for the single point displacement of 300 μm (at the center of the shaft) have been performed to simulate the impact hammer response of shafts. The generalized approximation of the coefficient of internal damping in bond graph

leads to an error in the value of damping ratio response. The experimental damping ratio differs from the simulated damping ratio by 15% for Al 6061, 12.31% for 4% Al_2O_3/Al 6061 composite, 10.08% for 8% Al_2O_3/Al 6061 composite, and 6.25% for 12% Al_2O_3/Al 6061 composite for the same generalized coefficient of internal damping (μ_i). Therefore, calibrated values of internal damping coefficient were introduced for which damping ratio have the same value for each composition of composite respectively.

The value of amplitudes and natural frequencies were in proximity to the experimental results. The lowest amplitude was achieved for 12% Al_2O_3/Al 6061 composite shaft, both by experiment (52.4 μm) and by bong graph simulation (49.9 μm). Moreover, the highest natural frequency was also attained for 12% Al_2O_3/Al 6061 composite shaft in both cases with minor differences in values. Therefore, the strengthening effect of the reinforcement in composite has been confirmed.

Further, the dynamic response for MMC rotating shafts was simulated by taking the effective parameters through calibrated and mean-field homogenized values. The shaft speed has been entrained near to the corresponding modal natural frequencies. The orbit response was obtained at 200 Hz for first mode (approximately twice the natural frequency of the first mode) and 800 Hz (approximately twice the natural frequency of the second mode). The maximum first mode horizontal and vertical displacements were reduced by 0.44 mm and 0.26 mm due to an addition of reinforcement. Also, the maximum second mode horizontal and vertical displacements were reduced by 0.310 mm and 0.18 mm due to the addition of reinforcement. This trend may have been different for longer shafts which can be taken as future scope of the present work.

References

Abaimov SG, Trofimov A, Sergeichev IV, Akhatov IS (2019) Multi-step homogenization in the Mori-Tanaka-Benveniste theory. Compos Struct. https://doi.org/10.1016/j.compstruct.2019.03.073

Allien VJ, Kumar H, Desai V (2019) Dynamic analysis and optimization of SiC reinforced Al6082 and Al7075 MMCs. Mater Res Express. https://doi.org/10.1088/2053-1591/ab038e

Assar A, El-Mahallawy N, Taha MA (1982) Influence of processing variables on structure and porosity in stir-casting Al–Cu alloys. Metals Technol 9(1):165–170. https://doi.org/10.1179/030716982803285990

Bajpai RP, Chandrasekhar U, Arankalle A (2018) Innovative design, analysis and development practices in aero-space and automotive engineering. Springer Singapore. https://doi.org/10.1007/978-81-322-1871-5

Banerjee JR, Su H (2006) Dynamic stiffness formulation and free vibration analysis of a spinning composite beam. Comput Struct 84:1208

Beaman JJ, Paynter HM (1993) Modeling of physical systems, pp 3.69–3.71

Benveniste Y (1987) A new approach to the application of Mori–Tanaka's theory in composite materials. Mech Mater 60:147–157

Bharti SK, Bisoi A, Sinha A, Samantaray AK, Bhattacharyya R (2019) Sommerfeld effect at forward and backward critical speeds in a rigid rotor shaft system with anisotropic supports. J Sound Vib 442:330–349. https://doi.org/10.1016/j.jsv.2018.11.002

Chang CY, Chang MY, Huang JH (2004a) Vibration analysis of rotating composite shafts containing randomly oriented reinforcements. Compos Struct 63:21

Chang MY, Chen JK, Chang CY (2004b) A simple spinning laminated composite shaft model. Int J Solids Struct 41:637

Chaturvedi SK, Tzeng GY (1991) Micromechanical modeling of material damping in discontinuous fiber three-phase polymer composites. Compos Eng 1(1):49–60

Chawla N, Shen Y (2001) Mechanical behavior of particle reinforced metal matrix composites. Adv Eng Mater 3:357–370. https://doi.org/10.1002/1527-2648(200106)3:6<357:AID-ADEM357>3.0.CO;2-I

Christensen RM, Malls WF (1972) Effective stiffness of randomly oriented fiber composites. J Compos Mater 6:518–532

Eshelby JD (1957) The determination of the elastic field of an ellipsoidal inclusion, and related problems. Proc Roy Soc A241:376–396

Guden M, Hall IW (1998) Dynamic properties of metal matrix composites: a comparative study. Mater Sci Eng A242:141–152

Halpin JC, Jerine K, Whitney JM (1971) The laminate analogy for two and three-dimensional composite materials. J Compos Mater 5:36–49

Hashim J, Looney L, Hashmi MSJ (2002) Particle distribution in cast metal matrix composites—Part II. J Mater Process Technol 123:258–263. https://doi.org/10.1016/S0924-0136(02)00099-7

Hubbard M (1979) Whirl dynamics of pendulous flywheels using bond graphs. J Franklin Instit 308(4):405

Hunt WH, Miracle DB (2001) Automotive applications of metal matrix composites, ASM handbook. In: Donaldson SL (ed) Composites 21. ASM International, Materials Park, pp 1029–1032

Imani SM, Goudarzi AM, Rabiee SM, Dardel M (2018) The modified Mori-Tanaka scheme for the prediction of the effective elastic properties of highly porous ceramics. Ceram Int 44:16489–16497. https://doi.org/10.1016/j.ceramint.2018.06.066

Jain AK, Rastogi V, Agrawal AK (2018) A novel approach to study effects of asymmetric stiffness on parametric instabilities of multi-rotor-system. J Sound Vib 413:159–172. https://doi.org/10.1016/j.jsv.2017.10.027

Kafi HR, Hosseini SAA (2019) Dynamic analysis of nonlinear rotating composite shafts excited by non-ideal energy source. ZAMM – J Appl Math Mech/Zeitschrift Für Angew Math Und Mech: e201800279. https://doi.org/10.1002/zamm.201800279

Kamarian S, Shakeri M, Yas MH (2018) Natural frequency analysis and optimal design of CNT/fiber/polymer hybrid composites plates using Mori-Tanaka approach, GDQ technique, and firefly algorithm. Polym Compos 39:1433–1446. https://doi.org/10.1002/pc.24083

Karnopp D (1981) Using bond graphs in nonlinear simulation problems. Simulation 36(6):183–192. https://doi.org/10.1177/003754978103600603

Karthikeyan M, Bisoi A, Samantaray AK, Bhattacharyya R (2015) Sommerfeld effect characterization in rotors with non-ideal drive from ideal drive response and power balance. Mech Mach Theory 91:269–288. https://doi.org/10.1016/j.mechmachtheory.2015.04.016

Koczak MJ, Khatri SC, Allison JE, Bader MG (2013) Metal-matrix composites for ground vehicle, aerospace, and industrial applications. In: Fundamentals of metal-matrix composites. Butterworth-Heinemann, Stoneham, pp 297–326

Lesieutre GA (1993) Damped structural composite material using resistively-shunted piezoelectric ceramic short fibers. In: Structural dynamics and materials conference, part 6, La Jolla, CA, USA, pp 3238–3243

Loewy RG, Piarulli VJ (1969) Dynamics of rotating shafts. The Shock and Vibration information Center, U.S. Department of Defense, Washington, DC, pp 6–10

Margolis DL (1980) Bond-graphs for some classic dynamic systems. Simulation 35(3):81–87. https://doi.org/10.1177/003754978003500302

Miracle DB (2005) Metal matrix composites—from science to technological significance. Compos Sci Technol 65(15–16):2526–2540

Mori T, Tanaka K (1973) Average stress in matrix and average energy of materials with misfitting inclusions. Acta Mettall 21:571–574

Mukherjee A, Samantaray AK (2000) SYMBOLS-2000 user's manual, India. High-Tech Consultants, STEP Indian Institute of Technology, Kharagpur, p 1

Mura T. Micromechanics of defects in solids. Springer Science & Business Media/Martinus Nijhoff, Dordrecht, 1987; 3, 580, p. 21

Naher S, Brabazon D, Looney L (2003) Simulation of the stir casting process. J Mater Process Technol 144:567–571. https://doi.org/10.1016/S0924-0136(03)00368-6

Oliveira L, Maia NMM, Marto AG, Silva RGA, Afonso FJ, Suleman A (2016) Modal characterization of composite flat plate models using piezoelectric transducers. Mech Syst Signal Process 79:16–29. https://doi.org/10.1016/j.ymssp.2016.02.046

Rastogi V, Mukherjee A, Dasgupta A (2011) Dynamic modeling of rotor shaft with internal damping driven through a dissipative coupling. Int J Model Simul Sci Comput 02:105–129. https://doi.org/10.1142/s1793962311000396

Ravi KR, Sreekumar VM, Pillai RM (2007) Optimization of mixing parameters through a water model for metal matrix composites synthesis. Mater Design 28:871–881. https://doi.org/10.1016/j.matdes.2005.10.007

Rawal SP (2001) Metal-matrix composites for space applications. J Miner Metals Mater Soc 53:14. https://doi.org/10.1007/s11837-001-0139-z

Sahu MK, Sahu RK (2017) Optimization of stirring parameters using CFD simulations for HAMCs synthesis by stir casting process. Trans Indian Inst Metals 70:2563–2570. https://doi.org/10.1007/s12666-017-1119-5

Sahu MK, Sahu RK (2018) Fabrication of aluminum matrix composites by stir casting technique and stirring process parameters. Optimization. https://doi.org/10.5772/intechopen.73485

Sarmasti AB, Yazdanirad M, Khezrabad M, Karbalaie M (2011) Effect of alumina particle size and thermal condition of casting on microstructure and mechanical properties of stir cast Al–Al2O3 composites. Mater Sci Technol 27(11):1653–1656. https://doi.org/10.1179/1743284710Y.0000000020

Shercliff HR, Ashby M (1994) Design with metal matrix composites. Mater Sci Technol 10(6):443–451. https://doi.org/10.1179/mst.1994.10.6.443

Singh S, Singh I, Dvivedi A (2017) Design and development of novel cost effective casting route for production of metal matrix. Int J Cast Metals Res 30(6):356–364. https://doi.org/10.1080/13640461.2017.1323605

Sino R, Baranger TN, Chatelet E, Jacquet G (2008) Dynamic analysis of a rotating composite shaft. Compos Sci Technol 68:337

Sobhaniaragh B, Batra RC, Mansur WJ, Peters FC (2017) Thermal response of ceramic matrix nanocomposite cylindrical shells using Eshelby-Mori-Tanaka homogenization scheme. Compos Part B Eng 118:41–53. https://doi.org/10.1016/j.compositesb.2017.02.032

Souza LF, Vandepitte D, Tita V, Medeiros R (2019) Dynamic response of laminated composites using design of experiments: an experimental and numerical study. Mech Syst Signal Process 115:82–101. https://doi.org/10.1016/j.ymssp.2018.05.022

Tandon GP, Weng GJ (1986) Average stress in the matrix and effective moduli of randomly oriented composites. Compos Sci Technol 27:111–132. https://doi.org/10.1016/0266-3538(86)90067-9

Tran VP, Brisard S, Guilleminot J, Sab K (2018) Mori–Tanaka estimates of the effective elastic properties of stress-gradient composites. Int J Solids Struct 146:55–68. https://doi.org/10.1016/j.ijsolstr.2018.03.020

Vaidya UK, Dadzie PE, Haque A, Mahfuz H, Jeelani S (1997) Nondestructive evaluation and characterization of microfibers modified textile carbon–phenolic and carbon–carbon composites. J Reinforced Plast Compos 16(11):968–1001

View publication stats

Journal of Applied Electrochemistry
https://doi.org/10.1007/s10800-024-02068-1

RESEARCH ARTICLE

CeO_2 and PEDOT:PSS modified conducting paper for organophosphate pesticide detection

Saroj Paneru[1] · Sweety[1] · Devendra Kumar[1]

Received: 15 October 2023 / Accepted: 8 January 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract

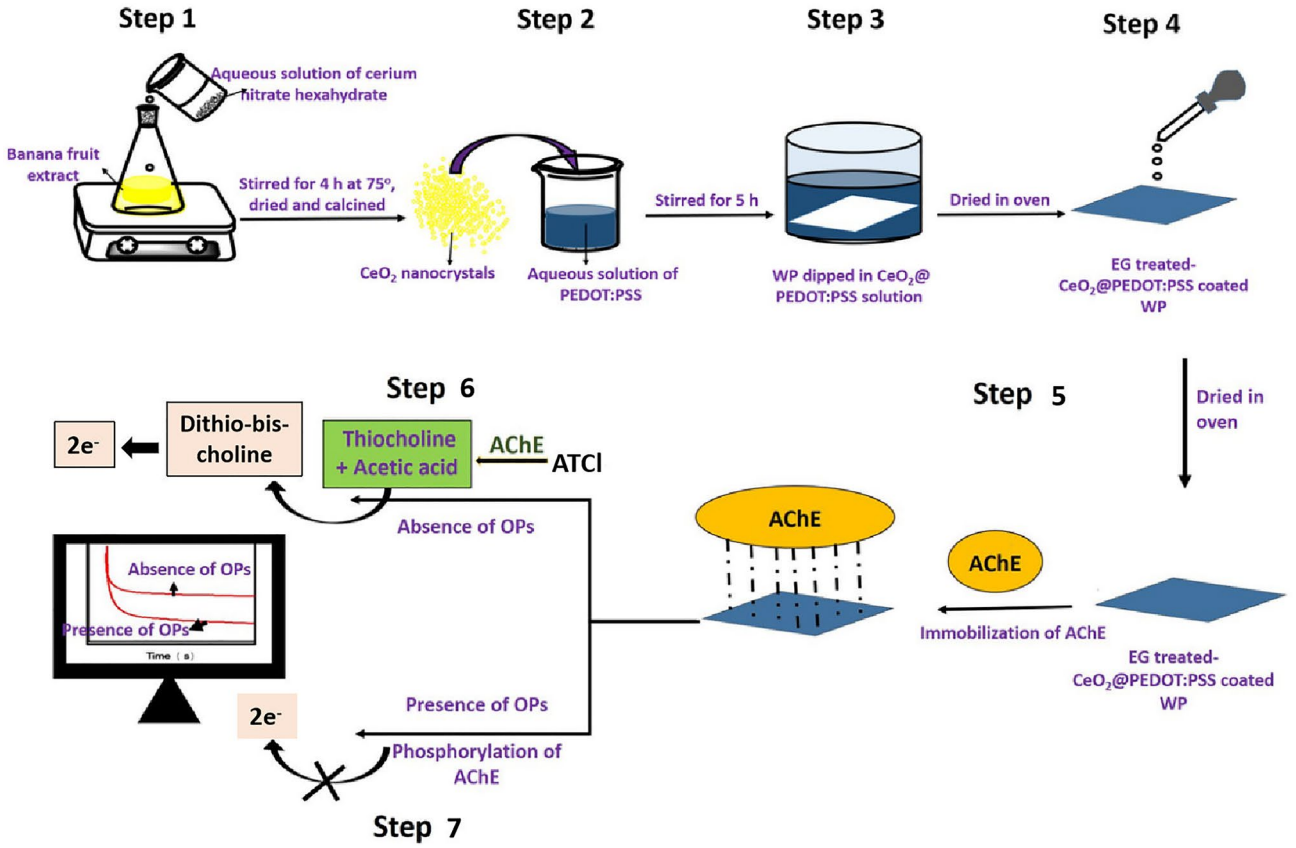
In this study, a novel, cost-effective, and environmental-friendly acetylcholinesterase (AChE)-based electrochemical biosensor was developed and utilized for paraoxon-ethyl (PE) detection. In this biosensor, cerium oxide (CeO_2) and poly(3,4 ethylenedioxythiophene):poly(styrenesulfonate) (PESOT:PSS) modified paper electrode was fabricated by a simple dip-coating method. The synergistic effect of CeO_2 and PEDOT:PSS enhance the performance of AChE-based biosensor in terms of sensitivity and selectivity. The electrical conductivity and stability of the fabricated electrode are further improved with the treatment of ethylene glycol (EG). EG decreases the coulombic interactions between PEDOT and PSS molecules, causes polymer chain reorientation, and improves the non-covalent interaction between cellulose molecules and PEDOT, which further leads to higher conductivity. All electrochemical results validate that the biosensor shows good selectivity ($K_{selectivity} < 1$), high sensitivity of 52.10 µA/nM, and a low detection limit of 0.12 nM. The accuracy and applicability of the proposed biosensor were also checked with two real samples. All results validate that the fabricated paper electrode can be a promising alternative to other conventional electrodes.

✉ Devendra Kumar
dkumar@dce.ac.in

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

Published online: 06 February 2024

Graphical abstract



Keywords Paper electrode · Acetylcholinesterase enzyme · PEDOT:PSS · CeO_2 · Pesticides

1 Introduction

Organophosphate pesticides (OPs) are extensively used to enhance productivity in the agriculture field due to their high insecticidal property and potency for pest control. However, their careless use on a large scale can lead to severe health issues, making it a global concern [1]. Pesticides leave behind residues in agricultural resources, soil, and water, which can enter the human body and lead to serious health problems, such as Alzheimer's, Parkinson's, cognitive disorders and thyroid problems. [2]. Thus, there is an urgent need to develop a sensitive and effective technique for monitoring the concentration of these harmful substances to protect living beings from their injurious effects.

Various analytical methods have been proposed for this purpose, including liquid chromatography associated with mass spectrometry [3], High-performance liquid chromatography [4], Raman spectroscopy [5], and enzyme-linked immunosorbent assays (ELISAs) [6, 7], etc. Although these methods are precise and accurate, their complex sample pretreatment, off-site procedure, and high equipment cost limit their practical application. Therefore, there is a need for a new approach that can overcome these flaws and meet the future demands of OPs detection.

The acetylcholinesterase (AChE)-based electrochemical biosensors have become increasingly popular as an alternative to analytical techniques for detecting OPs due to their on-site convenience, low-cost instrumentation, excellent selectivity, and rapid analysis. Their sensitivity significantly depends on the efficiency of enzyme immobilization and the electron transfer rate between the enzyme and electrode surface. To address these critical factors, researchers have explored various transducers in their studies [8–10]. Previously reported transducers require costly electrodes, including indium tin oxide (ITO), glassy carbon electrode (GCE), gold and platinum-coated glass, etc. These electrodes are expensive and prone to breaking when bent. These limitations further pose challenges for designing and fabricating flexible, cost-effective, and environment-friendly electrodes.

The conducting paper (CP)-based electrodes have the potential to revolutionize the field of biosensor technology due to their high cost-effectiveness, environmentally

friendly, flexible, and disposable nature [11, 12]. Conducting polymers (CPs) are an attractive choice for creating paper-based devices due to their distinctive properties, including solution processability, biocompatibility, affordability, flexibility, and high conductivity. The partially filled π-electrons in CPs overlap and create a wave function that extends throughout the molecule, allowing for the movement of electrons throughout the structure and improving the material's conductivity [11]. Out of all different CPs, poly(3,4 ethylenedioxythiophene):pol y(styrenesulfonate) (PEDOT:PSS) has been evaluated as a highly promising option for the fabrication of CP due to its ability to be uniformly entrapped within or on the surface of the paper using a straightforward dip-coating method [13, 14]. Additionally, it exhibits excellent stability, which can maintain its conductivity over a long period.

The performance of PEDOT:PSS-based paper electrode in terms of sensitivity, electrochemical kinetics, and stability can be further improved by doping it with nanomaterials like metal oxides (CuO [15], CeO_2 [16], TiO_2 [17]), graphene oxide [18], etc. Among all different nanomaterials, CeO_2 exhibits high affinity with the phosphate group, thereby increasing its applicability for OPs detection [19]. Their high oxygen storage capacity and redox nature make them an effective catalyst for the electrocatalytic reaction in AChE-based biosensors [20]. As per the author's review, the existing literature reveals that CeO_2@PEDOT:PSS-based paper electrode has not been reported so far. Therefore it can be applied to develop a novel and highly efficient method for OPs detection.

By considering the applications of CeO_2 and PEDOT:PSS, we proposed an efficient, novel, green, and cost-effective method for OPs detection using CeO_2@ PEDOT:PSS-coated paper electrode. The electrode fabrication involves the coating of CeO_2@PEDOT:PSS onto Whatman paper using a simple dip-coating method. Further, solvent (ethylene glycol) treatment enhanced the fabricated electrode's electrical conductivity. The AChE/EG-treated CeO_2@PEDOT:PSS/WP biosensor exhibits high sensitivity, selectivity, and stability due to the following reasons: (a) The incorporation of CeO_2 into PEDOT:PSS enhances the environmental stability, catalytic power and surface area of the prepared composite. (b) The addition of CeO_2 increases the rate of hydrolysis of acetyl-thiocholine chloride (ATCl) (substrate) and phosphorylation of the AChE enzyme, leading to improved electron kinetics. (c) EG improves the conductivity and stability of fabricated electrode, further enhancing the biosensor's sensitivity. (d) This study uses a cross-linking agent for efficient immobilization of enzyme.

2 Experimental section

2.1 Materials

PEDOT:PSS, cerium nitrate hexahydrate ($Ce(NO_3 \cdot 6H_2O)$ (99%), paraoxon-ethyl (PE), Whatman filter paper (pore size = 110 mm), acetyl-thio-choline chloride (ATCl), acetylcholinesterase (C3389-2KU) (AChE), ethylene glycol (EG), glutereldehyde, acetone, hydrochloric acid (HCl), ethanol are procured from Sigma-Aldrich, India. Monobasic and dibasic potassium phosphate, potassium ferro and ferri cyanide, and potassium chloride (KCl) are procured from Fisher-Scientific, India. The real samples and banana fruit were purchased from the local market of Delhi. Milli-Q water (18.2 MΩ cm) was used for the preparation of all solutions.

2.2 Biosynthesis of CeO_2 nanocrystals

The CeO_2 nanocrystals were synthesized by banana fruit extract. The fruit was washed, dried, and crushed before undergoing the extraction process using water as a solvent. In this process, 10 g of dried banana powder was soaked in 50 mL of distilled water for 24 h. The solution was centrifuged for 1 h at 500 rpm, followed by filtration with Whatman paper. After obtaining the filtrate, it was diluted with DI water in a ratio of 1:4 (10 mL of filtrate diluted with 40 mL of DI water). Following this, an aqueous solution of cerium nitrate hexahydrate (2.17 g in 50 mL DI water) was added and stirred for 15 min. Then, 4 M NaOH was gradually introduced until the pH reached 10. The obtained solution was stirred for a duration of 4 h at a temperature of 75 °C. Subsequently, the solution was subjected to centrifugation and washed several times with DI water. The obtained white precipitate was subjected to drying at a temperature of 80 °C and calcined in a muffle furnace at 300 °C with a heating rate of 4 °C/min (Fig. 1, step 1).

2.3 Fabrication of PEDOT:PSS/WP, CeO_2@ PEDOT:PSS/WP, and EG-treated CeO_2@ PEDOT:PSS/WP conducting papers

The Whatman paper (WP) was utilized for the fabrication of conducting papers. Firstly, the WP underwent stabilization by immersion in a solution of tween-20 surfactant, followed by 15 min of sonication. Subsequently, the treated WP was dried in an oven at 100 °C before being utilized in the fabrication of electrodes.

For the fabrication of PEDOT:PSS-coated WP, the stabilized WP was immersed into the suspension of PEDOT:PSS (2 mL in 10 mL water) for 1 h followed by drying in an oven

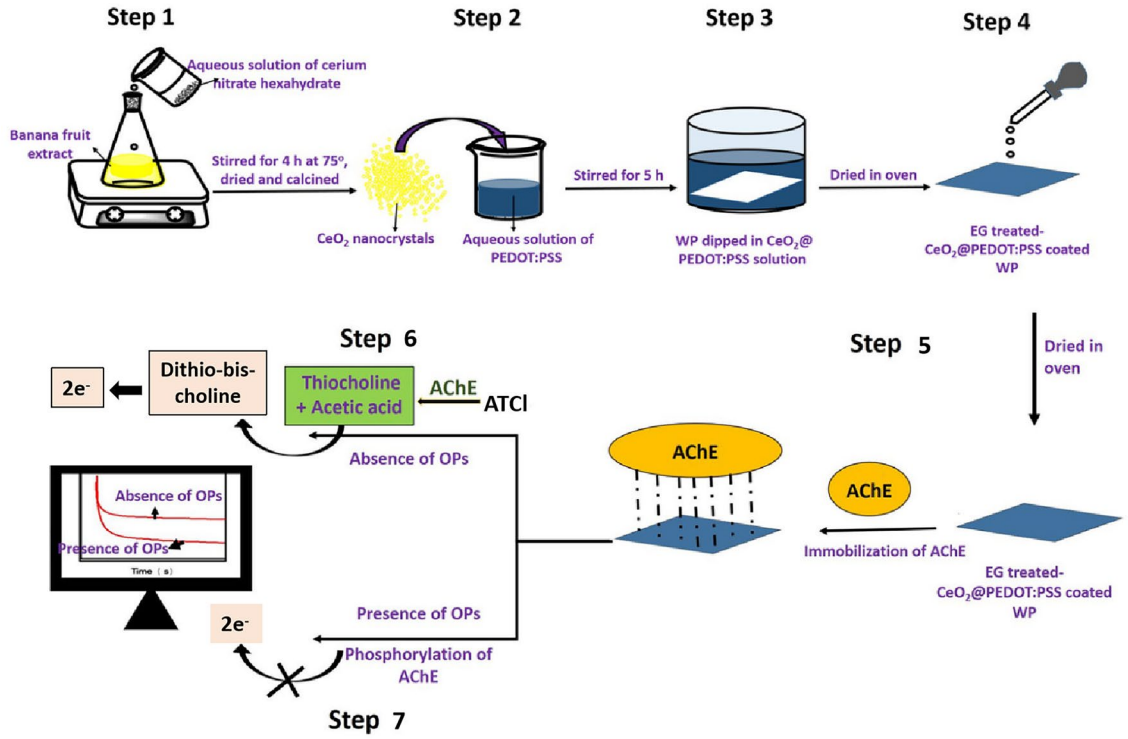


Fig. 1 Schematic diagram for the synthesis of CeO_2 nanocrystals, fabrication of AChE/EG-treated CeO_2@PEDOT:PSS/WP-based electrochemical biosensor, and mechanism for PE detection using the fabricated biosensor

at a temperature of 100 °C. To obtain better flexibility and conductivity of conducting paper, this process was repeated 4 times. The colour change from white to blue validated the successful coating of the polymer into WP.

The fabrication of CeO_2@PEDOT:PSS-coated WP was carried out by dissolving 5 mg CeO_2 nanocrystals into the aqueous solution of PEDOT:PSS through magnetic stirring for 5 h (Fig. 1, step 2). Then, the stabilized WP was immersed into the resultant solution of CeO_2 and PEDOT:PSS for 1 h followed by drying in an oven at 100 °C. This procedure was further repeated four times to fabricate CeO_2@PEDOT:PSS-coated WP (Fig. 1, step 3). For more stability and conductivity, the prepared electrode was subjected to a 37.5% ethylene glycol (EG) solution and further dried in an oven at 100 °C (Fig. 1, step 4).

2.4 Biosensor (AChE/EG-treated CeO_2@PEDOT:PSS/WP) fabrication

In the biosensor fabrication process, firstly, the conducting paper electrode undergoes incubation for 2 h at 25 °C with 0.2% glutaraldehyde (10 µL). Following this, 30 µL AChE enzyme (0.1 mg/mL) was drop-casted. Then, the fabricated biosensor was overnight kept in the incubation chamber at a temperature of 4 °C. The biosensor was further washed with phosphate buffer saline (PBS) before conducting the electrochemical analysis. The schematic representation of the proposed biosensor is illustrated in Fig. 1, step 5.

This study utilizes chemical cross-linking to immobilize the AChE enzyme, employing glutaraldehyde as the cross-linking agent. This process involves the formation of covalent bonds between the $-NH_2$ groups of the AChE enzyme and the −OH groups of PEDOT:PSS, facilitated by the aldehyde (−CHO) groups present in glutaraldehyde. Consequently, this establishes a stable and immobilized enzyme layer.

3 Results and discussion

3.1 Working mechanism of fabricated sensor for paraoxon-ethyl detection

The developed biosensor was used for PE detection by monitoring the generated current in both the presence and absence of PE. In the absence of PE, the AChE enzyme catalyzed the hydrolysis of ATCl. This reaction produced acetic acid and thiocholine as primary products. Thiocholine was subsequently oxidized, forming dithio-bis-choline and releasing two electrons, thereby generating a detectable speak current (Fig. 1, step 6). Conversely, the presence of PE triggers the phosphorylation of the AChE

enzyme, consequently inhibiting its catalytic activity in the hydrolysis of ATCl. This inhibition results in a reduction in the peak current intensity (Fig. 1, step 7).

3.2 Contribution of CeO_2 to the biosensor's performance

Recent studies reveal that Ce cations located on the surface of CeO_2 play a crucial role in activating the hydrolysis of esters and catalyzing the phosphorylation reaction with high efficiency [21]. In the current study, the CeO_2 increases the phosphorylation rate of AChE receptors through the Lewis acidity of Ce^{4+} ions. The Ce^{4+} ions are able to coordinate with the oxygen atoms in OPs, thereby facilitating the dissociation of the P-O bond (illustrated in Fig. 2A). Moreover, CeO_2 exhibits remarkable catalytic activity due to its redox properties. When cerium is reduced from Ce^{4+} to Ce^{3+}, oxygen vacancies are created, leading to an increase in the number of active catalytic sites on the surface of CeO_2 [20]. Figure 2B depicts how CeO_2 facilitates the rate of hydrolysis of ATCl. The CeO_2 acts as a catalyst due to its variable oxidation state, promoting the substitution of acetic acid and thiocholine, enhancing the reaction's electron kinetics.

3.3 Structural characterization

XRD and FTIR spectroscopy was used for the structural characterization of all fabricated electrodes and materials. Figure 3A illustrates the XRD diffraction pattern of PEDOT:PSS-coated-WP, CeO_2 nanocrystals and CeO_2 @ PEDOT:PSS-coated-WP through curve (a), (b) and (c), respectively. In curve (a), the peaks located at 16° and 23.2° are characteristic of PEDOT:PSS and identify their successful coating on the surface of WP [22, 23]. In curve (b), the characteristic peaks positioned at 29°, 33.5°, 48°, 56.8°, 70°, 77°, and 79.3° confirm that CeO_2 was successfully synthesized [24]. Finally, in curve (c), all the characteristic peaks of both PEDOT:PSS and CeO_2 are present, indicating both materials were successfully deposited into the surface of WP.

The FTIR spectra of CeO_2 (curve a), PEDOT:PSS/WP (curve b), and CeO_2 @PEDOT:PSS/WP are illustrated in Fig. 3B. In curve (a), the band appears at 712 cm^{-1}, corresponding to the Ce–O bond vibrations. While the bands located at 1059 and 1331 cm^{-1} are attributed to the vibrational modes of Ce–O–Ce bonds. The band appears at 1609 cm^{-1} due to bending vibrations of absorbed H_2O. All bands confirm the successful formation of CeO_2 [24]. In curve (b), the bands at 656 and 1054 cm^{-1} are attributed to the stretching vibrations of C–S and C–O in the thiophene

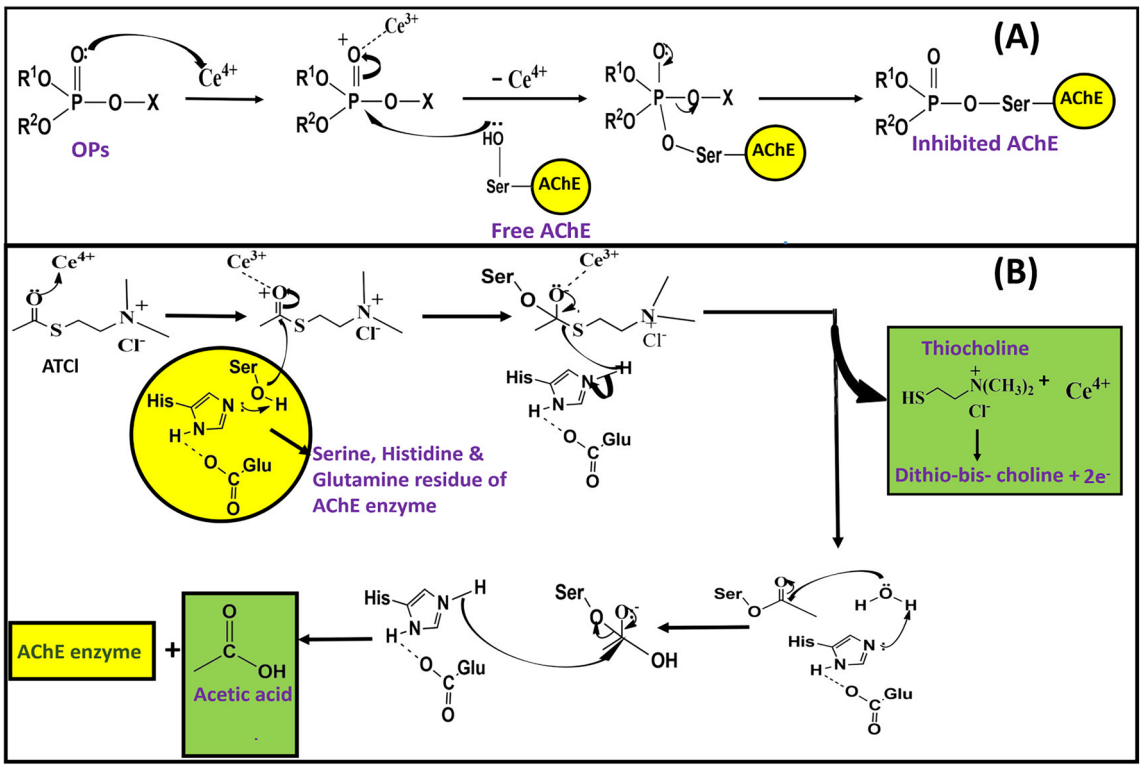


Fig. 2 The CeO_2 catalyzed **A** phosphorylation of AChE enzyme, and **B** electrocatalytic reaction of AChE, and ATCl

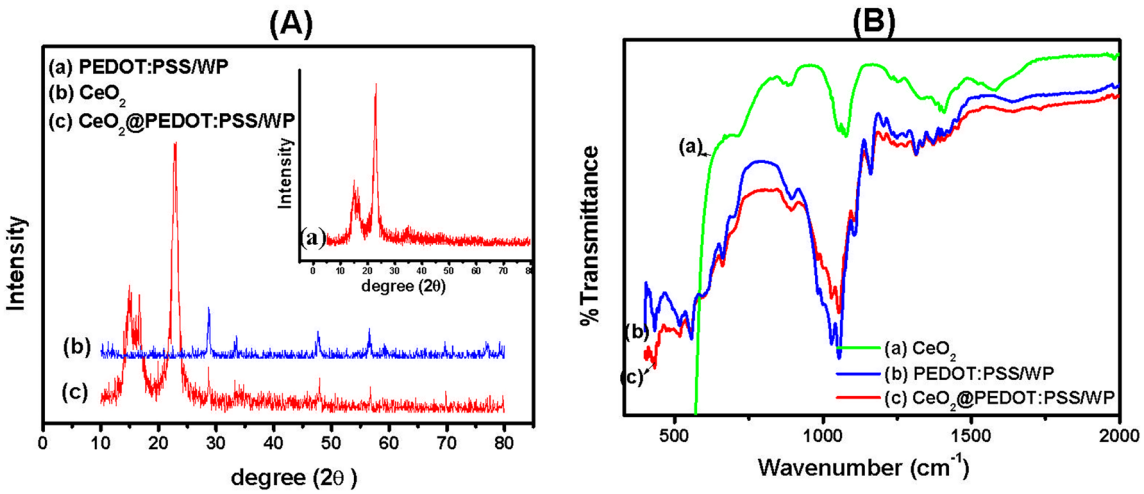


Fig. 3 **A** XRD pattern of (a) PEDOT:PSS/WP, (b) CeO_2 and (c) CeO_2@PEDOT:PSS/WP. **B** FTIR of (a) CeO_2, (b) PEDOT:PSS/WP and (c) CeO_2@PEDOT:PSS/WP

ring of PEDOT. The 1106 and 1162 cm^{-1} bands correspond to SO_3^- asymmetric stretching vibrations in PSS. The bands at 1516, and 1637 cm^{-1} are due to C–C and C=C stretching vibrations, respectively. These bands confirm the successful coating of PEDOT:PSS into the surface of WP. The FTIR spectra of CeO_2@PEDOT:PSS/WP (curve c) is almost similar to the FTIR spectra of PEDOT:PSS/WP. The characteristic vibrational bands of CeO_2 are not distinguished in FTIR spectra of CeO_2@PEDOT:PSS/WP due to very intensive bands of PEDOT:PSS [13, 14].

3.4 Morphological characterization

Figure 4A and B represent the SEM (scanning electron microscopy) and TEM (transmission electron microscopy) images of CeO_2 nanocrystals. The TEM image of CeO_2 confirms that the size of nanocrystals is less than 50 nm. The SEM image of bare WP is illustrated in Fig. 4C. In this figure, the fiber-like structure of cellulose is observed. Further, the SEM image of PEDOT:PSS/WP is illustrated in Fig. 4D. This figure clearly illustrates the uniform distribution of PEDOT:PSS over the surface of WP. In Fig. 4E, CeO_2 nanocrystals are clearly observed in PEDOT:PSS matrix, which confirms CeO_2@PEDOT:PSS is successfully coated into the WP surface. Moreover, Fig. 4F represents the morphological study of EG-treated CeO_2@PEDOT:PSS electrode.

Figure 5A and B illustrate the EDX (energy dispersive X-ray) spectra of CeO_2 nanocrystals and CeO_2@PEDOT:PSS/WP electrode, respectively. The presence of all peaks related to Ce and O in Fig. 5A confirms the successful formation of CeO_2 nanocrystal, while all peaks related to C, S, Ce, and O in Fig. 5B confirm the successful fabrication of CeO_2@PEDOT:PSS/WP electrode. The appearance of an additional peak in both figures is attributed to the gold coating applied to the samples for SEM-EDX analysis.

3.5 Electrical conductivity study of all fabricated electrodes

The electrical conductivity of all modified electrodes was studied by the Four-Probe technique. The WP coated with PEDOT:PSS shows a conductivity of 5.2×10^{-3} S/cm (Table 1). Further, the CeO_2 incorporated PEDOT:PSS/WP leads to an enhanced conductivity of 3.4×10^{-2} S/cm. Additionally, to improve the stability and conductivity, CeO_2@PEDOT:PSS/WP electrode was treated with different concentrations of ethylene glycol (EG), including 50, 37.5, 25, and 12.5% (V/V). The data in Table 1 highlights that the 37.5% EG treatment notably enhance electrode conductivity. Therefore, 37.5% EG was used for all subsequent studies. EG decreases the coulombic interactions between PEDOT and PSS molecules, causes the polymer chain reorientation and improves the non-covalent interaction between cellulose molecules and PEDOT, leading to higher conductivities.

3.6 Electrochemical studies of all fabricated electrodes

The chronoamperometry studies of PEDOT:PSS/WP (curve a), CeO_2@ PEDOT:PSS/WP (curve b), AChE/EG-treated CeO_2@PEDOT:PSS/WP (curve c), and EG-treated CeO_2@PEDOT:PSS/WP (curve d) electrodes are illustrated in Fig. 6A. The highest electrochemical current value for EG-treated CeO_2@PEDOT:PSS/WP electrode (1.0 mA) represents its highest conductivity compared to PEDOT:PSS/WP

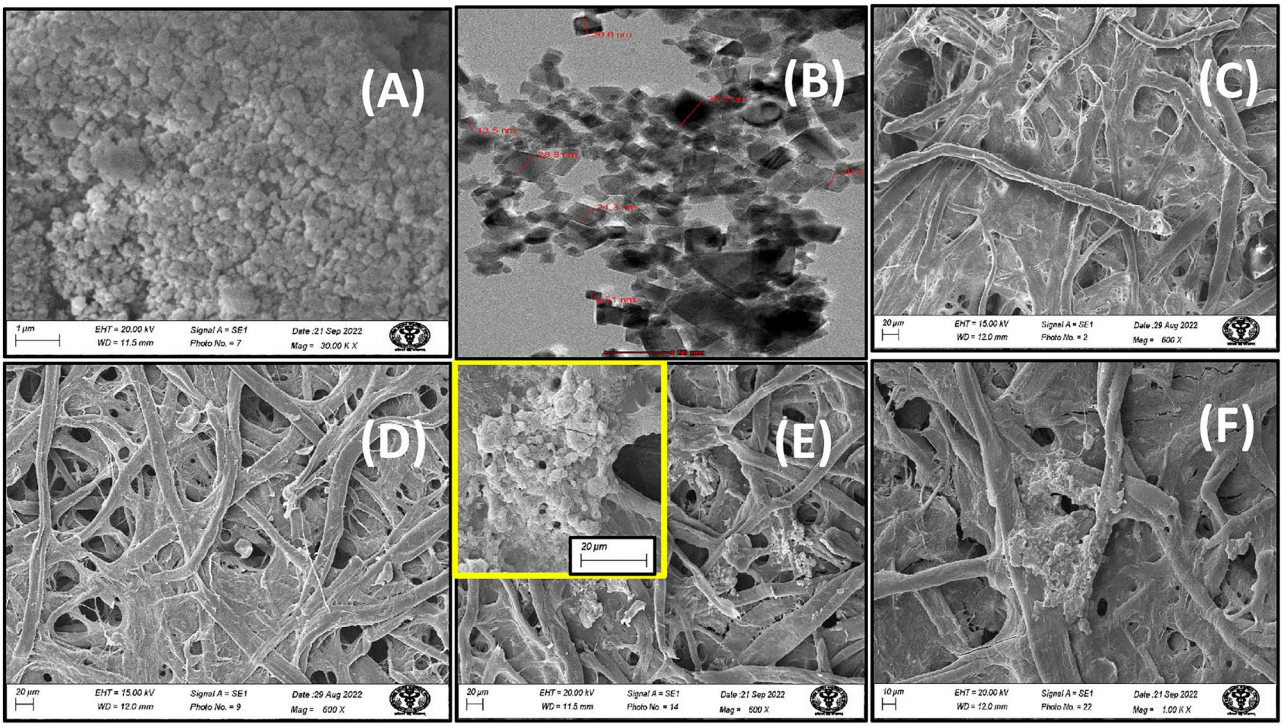


Fig. 4 **A** SEM image of CeO_2 nanocrystals, **B** TEM image of CeO_2 nanocrystals, **C** SEM image of WP, **D** SEM image of PEDOT:PSS/WP electrode, **E** SEM image of CeO_2@PEDOT:PSS/WP electrode, and **F** SEM image of EG-treated CeO_2@PEDOT:PSS/WP electrode

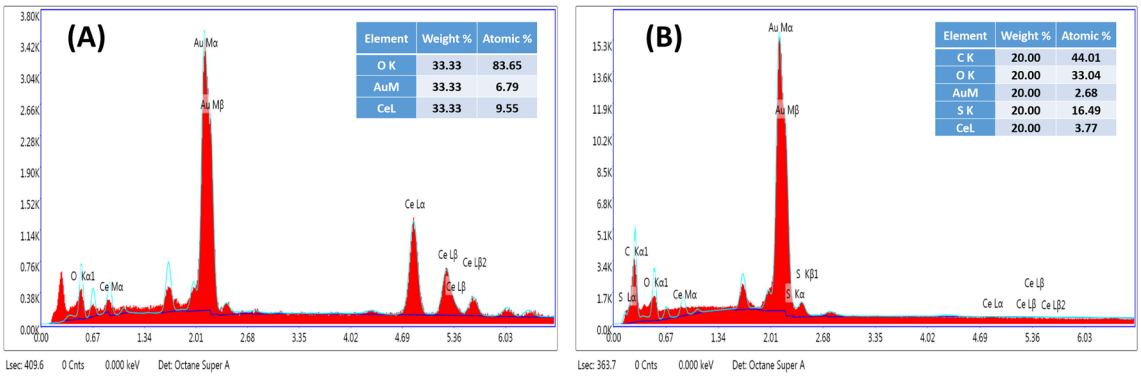


Fig. 5 SEM-EDX analysis of **A** CeO_2 and **B** CeO_2@PEDOT:PSS/WP electrode

Table 1 Electrical conductivity studies of all fabricated electrodes

S. No.	Modified electrodes	Conductivity (S/cm)
1.	PEDOT:PSS-coated-WP (PEDOT:PSS/WP)	5.2×10^{-3}
2.	CeO_2@ PEDOT:PSS/WP	3.4×10^{-2}
3.	50% EG-treated CeO_2@ PEDOT:PSS/WP	3.8×10^{-2}
4.	37.5% EG-treated CeO_2 @ PEDOT:PSS/WP	1.4×10^{-1}
5.	25% EG-treated CeO_2@ PEDOT:PSS/WP	6.5×10^{-2}
6.	12.5% EG-treated CeO_2@ PEDOT:PSS/WP	4.5×10^{-2}

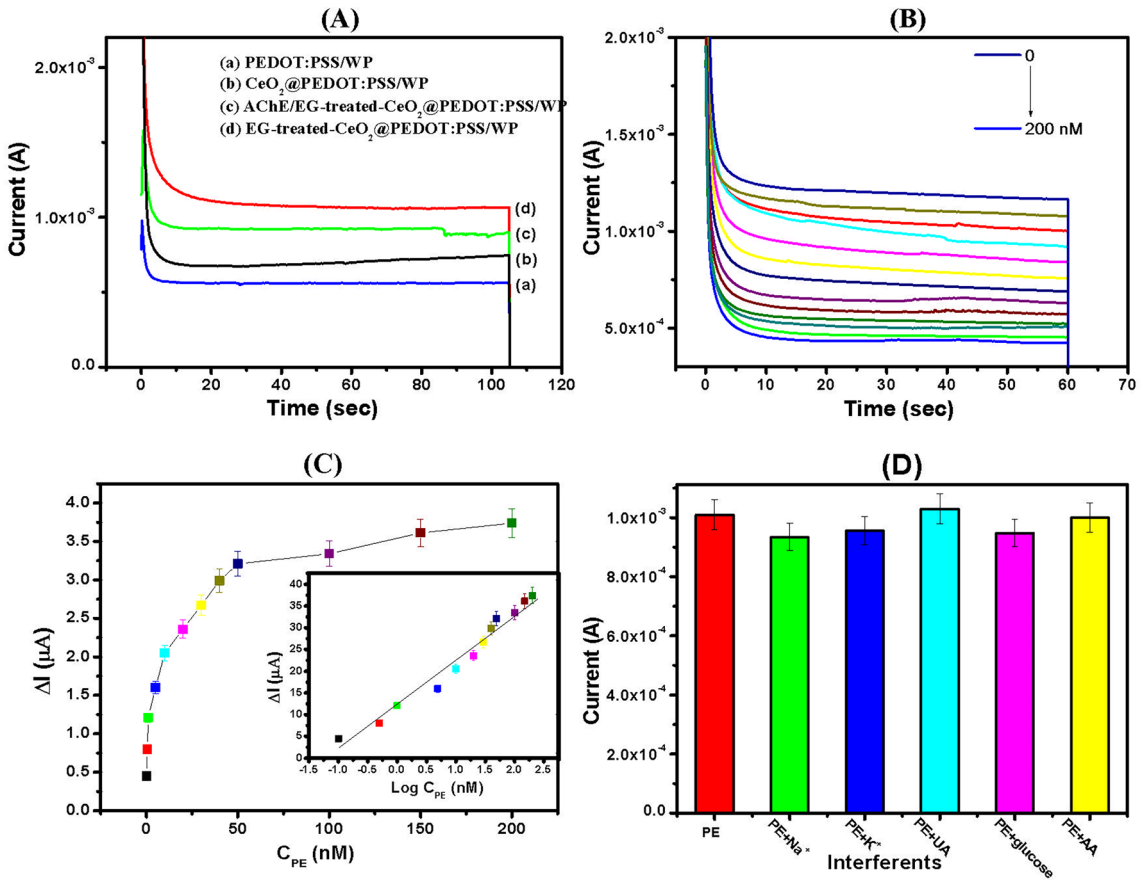


Fig. 6 **A** Chronoamperometry plot obtained for (a) PEDOT:PSS/WP, (b) CeO_2@PEDOT:PSS/WP, (c) AChE/EG-treated CeO_2@PEDOT:PSS/WP, and (d) EG-treated CeO_2@PEDOT:PSS/WP, **B** Electrochemical response study of AChE/EG-treated CeO_2@PEDOT:PSS/WP electrode after incubation of different concentrations (0, 0.1, 0.5, 1, 5, 10, 20, 30, 40, 50, 100, 150, and 200 nM) of PE. **C** Calibration plot of inhibition (ΔI) vs. concentration of PE (C_{PE}). Inset of 6(C) shows a linear curve of ΔI vs. log (C_{PE}). **D** Interference study of AChE/EG-treated CeO_2@PEDOT:PSS/WP electrode in the presence of Na^+, K^+, uric acid (UA), glucose, and ascorbic acid (AA)

(0.56 mA) and CeO_2@PEDOT:PSS/WP (0.70 mA) electrodes. The immobilization of the AChE enzyme onto the surface of the EG-treated CeO_2@PEDOT:PSS/WP electrode further causes the depletion of the electrochemical current. This decrease in current is attributed to the enzyme's non-conducting nature, which provides confirmation that the enzyme was effectively immobilized.

3.7 Optimization of experimental parameters

To optimize the bioelectrode response for PE detection, various parameters such as biosensor response time and pH of electrolytic solution were investigated using chronoamperometry. Initially, the response time for PE detection was explored using the AChE/EG-treated CeO_2@PEDOT:PSS/WP electrode, varying the time between 1 and 6 min, as depicted in Fig. S2A. The graph illustrates a consistent decline in current response until the 3 min mark, beyond which a plateau suggests saturation in the current. This indicates that maximum PE binding to the AChE enzyme occurred within this brief 3 min period, establishing 3 min as the optimal response time for PE detection in the presence of CeO_2.

While, in the absence of CeO_2, the response time extended to 8 min using the AChE/EG-treated PEDOT:PSS/WP electrode (Fig. S2B). This confirms that the inclusion of CeO_2 significantly enhances the rate of the electrocatalytic reaction between ATCl and the AChE enzyme, leading to a notable improvement in the response time for PE detection. Consequently, the AChE/EG-treated CeO_2@PEDOT:PSS/WP electrode was chosen for further studies.

The effect of pH on enzyme activity in the developed biosensor is also depicted in Fig. S2C. It is evident from the figure that the enzyme achieves its highest activity at a pH of 7, resulting in the highest current. Therefore, this optimal pH value was utilized in all subsequent biosensing experiments.

3.8 Electrochemical sensing studies

The biosensor's (AChE/EG-treated CeO_2@PEDOT:PSS/WP) performance was examined as a function of increasing concentration of PE (0.1–200 nM) using chronoamperometry in 7 pH PBS containing 3 mM acetyl thiocholine chloride (ATCl) and 5 mM $[Fe(CN)_6]^{3-/4-}$. Figure 6B demonstrated that the peak current decreases with an increase in the concentration of PE. The PE leads to the phosphorylation of AChE receptors, resulting in a decrease in the hydrolysis rate of ATCl and ultimately leading to a reduction in the peak current. Figure 6C represents the celebration plot of inhibition (ΔI) vs. PE concentration (C_{PE}). The inset of Fig. 6C shows a linear relationship between ΔI and the logarithm of PE concentration and is defined by following the regression equation:

$$\Delta I = 52.10\,\mu\text{A/nM}\log C_{PE} + 11.97\,\mu\text{A};\ R = 0.9856. \quad (1)$$

The sensitivity of the proposed biosensor (AChE/EG-doped-CeO_2@PEDOT:PSS/WP) is calculated as 52.10 µA/nM, and this value is obtained using the slope of the regression equation mentioned above (Eq. 1). Furthermore, the limit of detection (LOD) for the proposed biosensor was determined to be 0.12 nM using Eq. (2) [25]:

$$\text{LOD} = 3\sigma/S. \quad (2)$$

Here, S is sensitivity and σ is the standard deviation of the blank electrode.

The performance of the proposed biosensor in OPs detection was further evaluated and compared to similar works that utilized AChE enzyme and non-enzymatic biosensor. The analysis revealed that the proposed biosensor outperforms other considered biosensors in terms of three metrics, including sensitivity, LOD, and linear range (refer to Table 2). Additionally, this proposed conducting paper-based biosensor provides a straightforward, cost-effective, and environmentally friendly approach of detecting multiple OPs, marking a significant advancement over conventional methods.

3.9 Selectivity, reproducibility, and stability studies

To study the selectivity of AChE/EG-treated CeO_2@PEDOT:PSS/WP-based biosensor for PE detection, an interference study was performed in the presence of various interfering species, including Na^+, K^+, uric acid (UA), glucose, and ascorbic acid (AA). To conduct this study, 10 nM of specified interferents are added with the coexistence of 10 nM of PE. The bar diagram (Fig. 6D) demonstrated that the current response remains relatively unchanged with the addition of the interfering species, indicating excellent selectivity of the proposed biosensor for PE detection.

The $K_{selectivity}$ (selectivity coefficient) value for each interfering species was calculated by Eq. (3) and was found to be less than 1, indicating the biosensor's performance was minimally affected by the considered interferents [26].

$$K_{\text{selectivity}} = I_{\text{interferent}}/I_{PE}. \quad (3)$$

Now, to evaluate the stability of the proposed biosensor, the chronoamperometry response was analyzed every 5 days for a total of 25 days, when the biosensor was stored at 4 °C. After 25 days the current response remained 92.6% of its initial value, which indicated good stability of the developed biosensor (Fig. S3A).

Finally, to check the reproducibility of the developed biosensor, the chronoamperometry response of five different electrodes was analyzed (Fig. S3B). The results indicate that the electrodes' peak current does not differ significantly, with a low relative standard deviation (RSD) of 1.44%, indicating excellent reproducibility.

3.10 Real sample analysis

The precision and applicability of the proposed biosensor were evaluated by measuring the concentration of PE in two

Table 2 Comparative study of AChE/EG-treated CeO_2@ PEDOT:PSS/WP biosensor with recently developed biosensors of PE detection

S. No.	Fabricated electrodes	Linear range	Detection limit	Sensitivity	Sensor type	References
1.	AChE/MWCNTs-CS/AuNPs/SPCE	0.01–100 µM	0.03 µM	46.20 µA(µM)$^{-1}$	Enzymatic	[27]
2.	SPE/PHA/mPEG	1–10 µM	0.36 µM	0.05 µA(µM)$^{-1}$	Nonenzyme	[28]
3.	$BiVO_4$/flexible SPCE	NA	0.03 µM	0.345 µA(µM)$^{-1}$cm^{-2}	Nonenzyme	[29]
4.	GCE/YbM/f-CNFs	0.01–12 µM	0.002 µM	2.87 µA(µM)$^{-1}$ cm^{-2}	Nonenzyme	[30]
5.	Gr@NiFeSP	0.012–10 µM	3.7 nM	10.24 µA(µM)$^{-1}$	Nonenzyme	[31]
6.	Zn–Ni–P–S/GO/GCE	1–200 µM	35 nM	0.06369 µA(µM) $^{-1}$	Nonenzyme	[32]
7.	AChE/Au-Pt-PDA	0.5–1000 nM	0.85 nM	NA	Enzymatic	[33]
8.	PdNPs/PNIPAM-CT microgel/ GCE	0.01 µM to 1.3 mM	0.7 nM	1.60 µA(µM)$^{-1}$ cm^{-2}	Nonenzyme	[34]
9.	AChE/EG-doped-CeO $_2$ @PEDOT:PSS /ITO	0.1–200 nM	0.12 nM	52.10 µA(nM) $^{-1}$	Enzymatic	Present work

Table 3 PE detection in soil and carrot samples

Sample	Added amount (nM)	Found amount (nM)	Recovery (%)	RSD (%)
Soil	0.1	0.09	93	4.41
	0.5	0.49	97	2.12
	1	0.98	98	0.74
	5	4.91	98	1.12
	10	9.45	94	4.06
Carrot	0.1	0.09	94	3.78
	0.5	0.51	103	2.71
	1	1.07	107	5.07
	5	5.21	104	2.90
	10	9.87	98	0.73

real samples. In order to prepare the real samples for analysis, the necessary samples (soil and carrot) were collected and mixed with the paste of each one in a 7 pH PBS buffer. Then, the mixture was subjected to sonication and centrifugation at 3500 rpm for 15 to 20 min. After completing the centrifugation step, the supernatant liquid was carefully separated and spiked with various concentrations of PE, including 0.1, 0.5, 1, 5, and 10 nM. The spiked mixture was then used for chronoamperometry analysis using the fabricated biosensor. Fig. S4A and B illustrate the chronoamperometry curves of PE spiked soil and carrot sample, respectively. The results presented in Table 3 clearly demonstrates the excellent performance of biosensor, with satisfactory recovery rates ranging from 93 to 104% and a good level of accuracy (as indicated by the related standard deviation (RSD) of less than 5%). These findings suggest that the biosensor is a reliable and effective tool for detecting and quantifying pesticides in food and environmental samples.

4 Conclusion

In this work, a CeO_2@PEDOT:PSS-based paper electrode was fabricated by dip-coating method and further subjected to an ethylene glycol solution for more stability and conductivity. The fabricated electrode was further used to develop an AChE-based electrochemical biosensor for PE detection. In order to improve the performance of the developed biosensor, an optimized process was conducted focused on two key parameters: pH and incubation time. The biosensor shows high sensitivity, selectivity and stability. The biosensor's performance was further evaluated by comparing its LOD, linear range and sensitivity with recently developed biosensors, which shows the outperformance of the proposed biosensor. The accuracy and applicability of the proposed biosensor in the real-world scenario were also checked. All results support the conclusion that this research offers a reliable, cost-effective, and sensitive method for OPs detection.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10800-024-02068-1.

Acknowledgements The first author acknowledges to the University Grants Commission (UGC), Government of India, New Delhi for their financial support (Grant No. 351/CSIR-UGC NET DEC.2016).

Author contributions SP: Conceptualization, methodology, validation, and writing original draft. S: Methodology, writing original draft. DK: Supervision, writing—review and editing.

Funding Not applicable.

Data availability Not applicable.

Declarations

Competing interest The authors declare no competing interests.

References

1. Rahmani T, Bagheri H, Behbahani M, Hajian A, Afkhami A (2018) Modified 3D graphene-Au as a novel sensing layer for direct and sensitive electrochemical determination of carbaryl pesticide in fruit, vegetable, and water samples. Food Anal Methods 11(11):3005–3014
2. Alavinia SJ, Mirvaghefi A, Farahmand H, Rafiee G, Alavinia SJ, Shiry N, Moodi S (2019) DNA damage, acetylcholinesterase activity, and hematological responses in rainbow trout exposed to the organophosphate malathion. Ecotoxicol Environ Saf 182:109311
3. Shin Y, Lee J, Lee J, Lee J, Kim E, Liu KH, Lee HS, Kim JH (2018) Validation of a multiresidue analysis method for 379 pesticides in human serum using liquid chromatography–tandem mass spectrometry. J Agric Food Chem 66(13):3550–3560
4. Xu ZL, Deng H, Deng XF, Yang JY, Jiang YM, Zeng DP, Huang F, Shen YD, Lei HT, Wang H, Sun YM (2012) Monitoring of organophosphorus pesticides in vegetables using monoclonal antibody-based direct competitive ELISA followed by HPLC–MS/MS. Food Chem 131(4):1569–1576
5. Liu B, Zhou P, Liu X, Sun X, Li H, Lin M (2013) Detection of pesticides in fruits by surface-enhanced Raman spectroscopy coupled with gold nanostructures. Food Bioprocess Technol 6:710–718
6. Hua X, Yang J, Wang L, Fang Q, Zhang G, Liu F (2012) Development of an enzyme linked immunosorbent assay and an immunochromatographic assay for detection of organophosphorus pesticides in different agricultural products. PLoS One 7(12):e53099
7. Lee WY, Lee EK, Kim YJ, Park WC, Chung T, Lee YT (2006) Monoclonal antibody-based enzyme-linked immunosorbent assays for the detection of the organophosphorus insecticide isofenphos. Anal Chim Acta 557(1–2):169–178
8. Yang G, He Y, Zhao J, Chen S, Yuan R (2021) Ratiometric electrochemiluminescence biosensor based on ir nanorods and CdS quantum dots for the detection of organophosphorus pesticides. Sens Actuators B 341:130008
9. Jia L, Zhou Y, Wu K, Feng Q, Wang C, He P (2020) Acetylcholinesterase modified AuNPs-MoS2-rGO/PI flexible film biosensor: towards efficient fabrication and application in paraoxon detection. Bioelectrochemistry 131:107392

10. Itsoponpan T, Thanachayanont C, Hasin P (2021) Sponge-like CuInS2 microspheres on reduced graphene oxide as an electrocatalyst to construct an immobilized acetylcholinesterase electrochemical biosensor for chlorpyrifos detection in vegetables. Sens Actuators B 337:129775
11. Kumar S, Umar M, Saifi A, Kumar S, Augustine S, Srivastava S, Malhotra BD (2019) Electrochemical paper based cancer biosensor using iron oxide nanoparticles decorated PEDOT: PSS. Anal Chim Acta 1056:135–145
12. Paschoalino WJ, Kogikoski S Jr, Barragan JT, Giarola JF, Cantelli L, Rabelo TM, Pessanha TM, Kubota LT (2019) Emerging considerations for the future development of electrochemical paper-based analytical devices. ChemElectroChem 6(1):10–30
13. Kumar S, Kumar S, Srivastava S, Yadav BK, Lee SH, Sharma JG, Doval DC, Malhotra BD (2015) Reduced graphene oxide modified smart conducting paper for cancer biosensor. Biosens Bioelectron 73:114–122
14. Paul G, Verma S, Jalil O, Thakur D, Pandey CM, Kumar D (2021) PEDOT: PSS-grafted graphene oxide-titanium dioxide nanohybrid-based conducting paper for glucose detection. Polym Adv Technol 32(4):1774–1782
15. Zhang Y, Liu Y, Su L, Zhang Z, Huo D, Hou C, Lei Y (2014) CuO nanowires based sensitive and selective non-enzymatic glucose detection. Sens Actuators B 191:86–93
16. Xie Y, Tu X, Ma X, Fang Q, Liu G, Dai R, Qu F, Yu Y, Lu L, Huang X (2019) A CuO-CeO_2 composite prepared by calcination of a bimetallic metal-organic framework for use in an enzyme-free electrochemical inhibition assay for malathion. Microchim Acta 186:1–9
17. Cui HF, Wu WW, Li MM, Song X, Lv Y, Zhang TT (2018) A highly stable acetylcholinesterase biosensor based on chitosan-TiO_2-graphene nanocomposites for detection of organophosphate pesticides. Biosens Bioelectron 99:223–229
18. Li Y, Bai Y, Han G, Li M (2013) Porous-reduced graphene oxide for fabricating an amperometric acetylcholinesterase biosensor. Sens Actuators B 185:706–712
19. Dong J, Wang X, Qiao F, Liu P, Ai S (2013) Highly sensitive electrochemical stripping analysis of methyl parathion at MWCNTs–CeO_2–Au nanocomposite modified electrode. Sens Actuators B 186:774–780
20. Lee KM, Brito M, DeCoster J, Linskens K, Mehdi K, Lee WI, Kim E, Kim H, Kwon G, Nam CY, Kim T (2022) Influence of oxidizing and reducing pretreatment on the catalytic performance of CeO_2 for CO oxidation. Mol Catal 528:112465
21. Manto MJ, Xie P, Wang C (2017) Catalytic dephosphorylation using ceria nanocrystals. ACS Catal 7(3):1931–1938
22. Yoo D, Kim J, Kim JH (2014) Direct synthesis of highly conductive poly (3,4-ethylenedioxythiophene): poly (4-styrenesulfonate) (PEDOT: PSS)/graphene composites and their applications in energy harvesting systems. Nano Res 7:717–730
23. Paneru S, Kumar D (2024) CuS modified PEDOT: PSS grafted paper-based electrochemical immunosensor for EpCAM biomarker detection. Mater Chem Phys 313:128687
24. Jayakumar G, Irudayaraj AA, Raj AD (2019) Investigation on the synthesis and photocatalytic activity of activated carbon–cerium oxide (AC–CeO_2) nanocomposite. Appl Phys A 125(11):742
25. Paneru S, Kumar D (2023) A novel electrochemical biosensor based on polyaniline-embedded copper oxide nanoparticles for high-sensitive paraoxon-ethyl (PE) detection. applied biochemistry and biotechnology. Jan 26:1–8
26. Paneru S, Kumar D (2023) Ag-doped-CuO nanoparticles supported polyaniline (PANI) based novel electrochemical sensor for sensitive detection of paraoxon-ethyl in three real samples. Sens Actuators B 379:133270
27. Hua QT, Ruecha N, Hiruta Y, Citterio D (2019) Disposable electrochemical biosensor based on surface-modified screen-printed electrodes for organophosphorus pesticide analysis. Anal Methods 11(27):3439–3445
28. Sgobbi LF, Machado SA (2018) Functionalized polyacrylamide as an acetylcholinesterase-inspired biomimetic device for electrochemical sensing of organophosphorus pesticides. Biosens Bioelectron 100:290–297
29. Gopi PK, Ngo DB, Chen SM, Ravikumar CH, Surareungchai W (2022) High-performance electrochemical sensing of hazardous pesticide paraoxon using $BiVO_4$ nano dendrites equipped catalytic strips. Chemosphere 288:132511
30. Karthik R, Kumar JV, Chen SM, Kokulnathan T, Yang HY, Muthuraj V (2018) Design of novel ytterbium molybdate nanoflakes anchored carbon nanofibers: challenging sustainable catalyst for the detection and degradation of assassination weapon (paraoxon-ethyl). ACS Sustain Chem Eng 6(7):8615–8630
31. Aghaie A, Khanmohammadi A, Hajian A, Schmid U, Bagheri H (2019) Nonenzymatic electrochemical determination of paraoxon ethyl in water and fruits by graphene-based NiFe bimetallic phosphosulfide nanocomposite as a superior sensing layer. Food Anal Methods 12(7):1545–1555
32. Xiao B (2022) Electrochemical sensor based on bimetallic phosphosulfide Zn–Ni–P–S nanocomposite-reduced graphene oxide for determination of paraoxon ethyl in agriculture wastewater. Int J Electrochem Sci 17(220672):2
33. Wu Y, Jiao L, Xu W, Gu W, Zhu C, Du D, Lin Y (2019) Polydopamine-capped bimetallic AuPt hydrogels enable robust biosensor for organophosphorus pesticide detection. Small 15(17):1900632
34. Mutharani B, Ranganathan P, Chen SM, Karuppiah C (2019) Enzyme-free electrochemical detection of nanomolar levels of the organophosphorus pesticide paraoxon-ethyl by using a poly (N-isopropyl acrylamide)-chitosan microgel decorated with palladium nanoparticles. Microchim Acta 186(3):1–1

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

Research Article

I·M3
Institute of Materials, Minerals & Mining

Materials Science and Technology
2024, Vol. 40(1) 47–53
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/02670836231212615
journals.sagepub.com/home/mst
Sage

Chopped carbon fibre reinforcement in high-strength concrete: Flexural characteristics

Nitin Lamba[1], Ritu Raj[1] and Poonam Singh[2]

Abstract

In this study, the flexural behaviour of high-strength concrete was tested with chopped carbon fibres that were 30 mm long and added in amounts of 0.2%, 0.4%, 0.6%, 0.8%, and 1%. Thirty-six prisms were evaluated experimentally for flexural strength at 28 days using an ultrasonic pulse velocity test and a flexural testing machine. X-ray diffraction and linear regression analysis were carried out to characterise crystalline material and linear correlation amongst different mixes. Increasing the number of carbon fibres in high-strength concrete led to unexpected and innovative findings, such as lower pulse velocities and improved flexural characteristics up to an optimum dosage point. In SEM images of chopped carbon fibres, debonding was the most prevalent cause of failure.

Keywords

Flexural strength, linear regression analysis, X-ray diffraction, carbon fibres, high-strength concrete, SEM, EDX

Received: 8 December 2022; accepted: 17 October 2023

Introduction

Due to the fast-growing world's modernisation, there is high natural resource consumption with rapidly growing demands of urbanisation, industrialisation, etc., resulting in the dumping of scrap materials all over places, which is becoming a tedious task for humans to tackle this problematic situation. Hence it is highly essential to dispose of these scrap materials is utmost required. Hence the reuse of scrap materials could be a solution to save our mother earth and the environment from the harmful side effects of these scrap materials. In construction, several scrap materials are used worldwide, such as fly ash, silica fume, tire shreds, carbon fibres, etc. Innovative research is going on to develop high-strength concrete using scrap materials; this way, apart from attaining high strength, we are helping reuse processes. Research has been going on to understand the role of silica fume and fly ash in developing high-strength concrete.[1] Several studies have been carried out on the properties of carbon fibre reinforcement in cement-based materials, reinforcement effect, and flexural toughness in concrete.[2–4]

Moreover, some researchers have helped the construction field by enlightening the world with chopped fibres for attaining high-strength concrete.[5,6] Several researchers have benefited the construction industry by raising public awareness of incorporating chopped fibres to generate high-strength concrete.[5–11] In addition, research has been done to see if the performance of regular concrete may be enhanced by adding a variety of fibres to the mix.[12–16] According to the authors, there is relatively little research on using chopped carbon fibre fragments to achieve high-strength concrete.[6,13,17,18] High-strength concrete was developed by reinforcing chopped carbon fibres into the concrete matrix, and its unique and intriguing character made its usage in building relatively straightforward. So, the present research aimed to learn about the properties of chopped carbon fibre reinforcement to design high-strength concrete using waste products such as silica fume and fly ash. The flexural property of high-strength concrete with a novel aspect ratio and dosage values of chopped carbon fibres was chosen to offer originality to the study. So, a new investigation into enhanced high-strength concrete is required due to the exceptional and erratic outcomes of the ultrasonic pulse velocity test and the flexural strength test setup. Destructive and non-destructive testing, X-ray diffraction (XRD), and analytical analysis were used to

[1]Department of civil engineering, Delhi Technological University, Shahbad Daulatpur, Delhi, India
[2]Department of Applied chemistry, Delhi Technological University, Shahbad Daulatpur, Delhi, India

Corresponding author:
Ritu Raj, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi, 110042, India.
Email: rituraj@dtu.ac.in

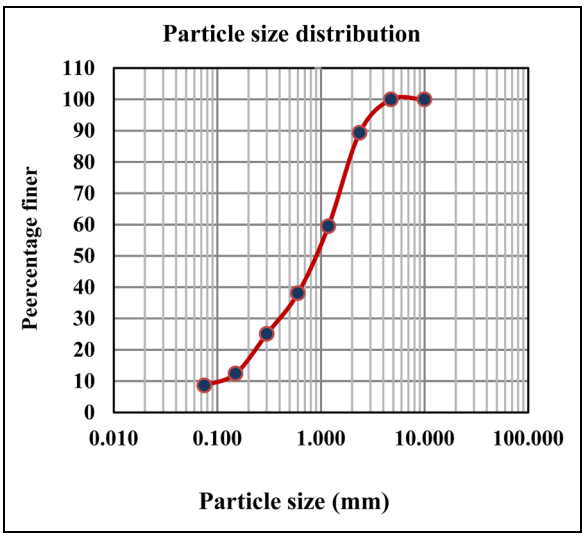


Figure 1. Particle size distribution curve.



Figure 2. Chopped carbon fibres.

Table 1. The outcomes of tests on the various physical properties of carbon fibre.

Physical properties	Result
Tensile strength (MPa)	3800
Fibre length (mm)	30
Bulk density ($g\ l^{-1}$)	555
Specific heat ($cal\ g^{-1}\ c^{-1}$)	0.15
Volume resistivity (cms)	2×10^3
Carbon content (%)	93
Elongation (%)	1.8

evaluate the feasibility of carbon fibre and the flexure performance of high-strength concrete in 36 prismatic beams of six different mixes, each with 30 mm average length and a dose value of 0.2%, 0.4%, 0.6%, 0.8%, and 1% of fibres.

Test methodology

Raw materials and concrete mix design

Fibre-reinforced high-strength concrete was created with the help of fly ash grade 1, silica fume, OPC 43 cement, coarse and fine aggregates, super plasticiser additive (S.P), carbon fibres, and water as chief ingredients. The preliminary cement tests were performed as per the IS 8112-2013 norm.[19] Ordinary Portland Cement, which has a minimum 43 MPa compressive strength, also known as OPC 43, was incorporated to develop concrete specimens of the desired strength. Specifications specified by 3812[20] were employed in developing concrete mix designs that included pozzolan material, that is, fly ash of grade 1 obtained from a local coal thermal power station. Preliminary experiments on this experiment were conducted in line with IS Code, AASHTO M307, and ASTM C-1240.[21–23] The mineral additive silica fume was employed for the concrete mix design. To meet the requirements of IS 9103:1999,[24] ACI WR-1010, a superplasticiser, which is a polycarboxylate ether-based chemical, was incorporated into the mix design as a water reducer; its optimal dose was determined with the aid of a marsh cone apparatus,[25] and it was fixed at 1.1%. An IS 383: 1970[26] sieve analysis test was carried out on zone II fine aggregates (F.A) to generate a particle size distribution curve for the experiment (see Figure 1). The tests were conducted using 10 mm nominal size coarse aggregates (C.A) to evaluate water absorption, specific gravity (S.G), and fineness modulus (F.M) results from preliminary tests. For sand, the values for water absorption, S.G, and F.M, were 1.2%, 2.6%, and 2.8%, respectively. According to the standards established by the IS codes,[27–29] the results for water absorption, S.G, and F.M for C.A were 7.3%, 2.81%, and 0.20%, respectively. Chopped fibre pieces of 30 mm length were incorporated; Figure 2 is a picture of the chopped carbon fibres used, and Table 1 lists the physical attributes of the tested carbon fibres. As a means of experimentation, Size $500 \times 100 \times 100$ mm prismatic beam samples were made with six types of mixes, including nominal mix and fibre additions of 0.2%, 0.4%, 0.6%, 0.8%, and 1% volume fraction (Vf). Once all the chief ingredients of the mix design were thoroughly mixed in a concrete mixer, transfer the prepared mix into the steel moulds; after thorough compaction of the mixture in the mould, the concrete had to cure for about 24 h before it was set. The samples were removed after 24 h from the steel moulds and kept in water for about 28 days, after which they were first tested for non-destructive testing (ultrasonic pulse velocity test). Then later flexural strength test was performed. The mix compositions achieved, which were used for sample preparation, are mentioned in Table 2.

Experimental test procedure

Ultrasonic pulse velocity test

Non-destructive tests were performed initially on the experimental samples prepared for the experiments, followed by destructive tests (four-point bending test). A unit of ultrasonic pulse velocity testing (UPVT) equipment,

Table 2. Mix composition for concrete mix design developed.

Sample types	Cement	Silica fume	Fly ash	C.A	F.A	S.P (%)	Water	Chopped fibre	
								Dosage (%)	Length (mm)
F1L0	442	41	106	1167	580	1.1	141	0	0
F2L30	442	41	106	1167	580	1.1	141	0.2	30
F3L30	442	41	106	1167	580	1.1	141	0.4	30
F4L30	442	41	106	1167	580	1.1	141	0.6	30
F5L30	442	41	106	1167	580	1.1	141	0.8	30
F6L30	442	41	106	1167	580	1.1	141	1	30

C.A: coarse aggregates; F.A: fine aggregates; S.P: super plasticiser additive.



Figure 3. Ultrasonic pulse velocity testing (UPVT) test setup.



Figure 4. Flexure test setup.

as depicted in Figure 3, was used in accordance with IS code[30] to conduct the tests which work on the technique of direct transmission. Below is the equation adopted to determine the pulse velocity of the samples prepared:

$$V = L / T \quad (1)$$

Pulse velocity (in km s^{-1}) for the cast prism beam samples needs to be estimated, L stands for the whole distance that separates the transducers, and T (in μ-s) is the time taken by pulse velocity to travel in between transducers.

Four-point bending test

Following the termination of the UPVT test, flexure strengths of prismatic beam samples of six various sample mixes of size $500 \times 100 \times 100$ mm were evaluated using a four-point loading method with flexural test equipment of 100 kN capacity applied at the mid-span of the sample in accordance with the standards set forth by ASTM C78.[31] Displacement-controlled four-point bending tests were conducted at a deflection rate of 0.6 mm min^{-1}. Below is the equation for determining flexural strength:

$$Fr = FL / bh^2 \quad (2)$$

In this equation, Fr denotes the flexure strength of the prismatic beam specimen being tested, L represents the span length of the sample being evaluated, b denotes the width of the specimen being tested that is equal to 50 mm, h denotes the depth of the sample being considered that is also equal to 50 mm, F denotes the ultimate load at which the specimen fails. The reading is taken from the flexure test apparatus in the manner depicted in Figure 4.

Result and discussion

Ultrasonic pulse velocity, flexural strength and X-ray diffraction analysis

Based on the findings acquired from the UPVT test setup, it was noted that the pulse velocity values decrease with an increase in Vf or dosage values of samples, as displayed in Figures 5 and 6. These findings are depicted in both of these figures. When compared to the specimen with no fibre content, it was discovered that sample F3L30 had the largest drop in pulse velocity values, which was approximately 39%. Amongst all the reinforced specimens, the decline in pulse velocity that was discovered to be the highest was around 11% in sample F3L30. This was compared to sample F2L30, which had a minor fibre content prepared. The pulse velocity values of the test specimens

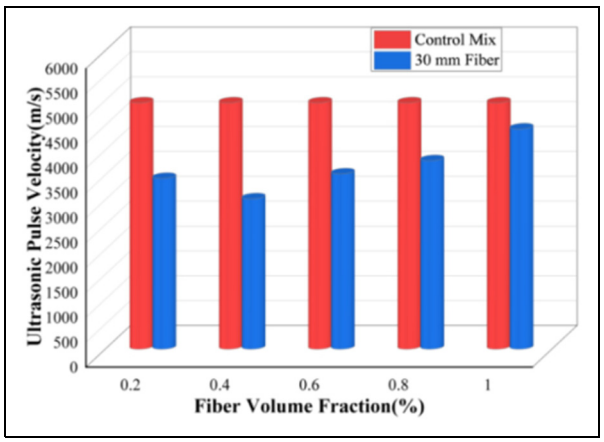


Figure 5. Fibre volume fraction versus ultrasonic pulse velocity testing (UPVT).

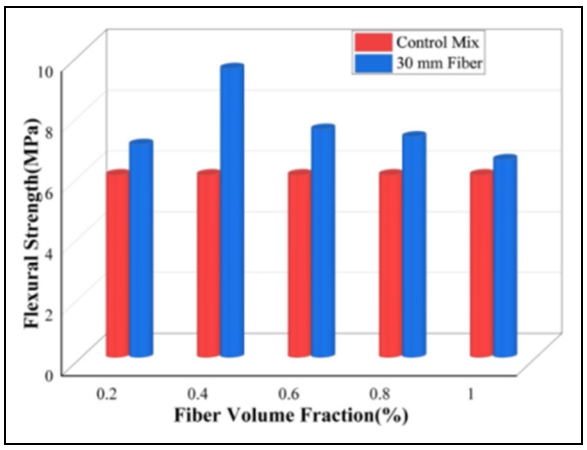


Figure 7. Fiber volume fraction versus Fr.

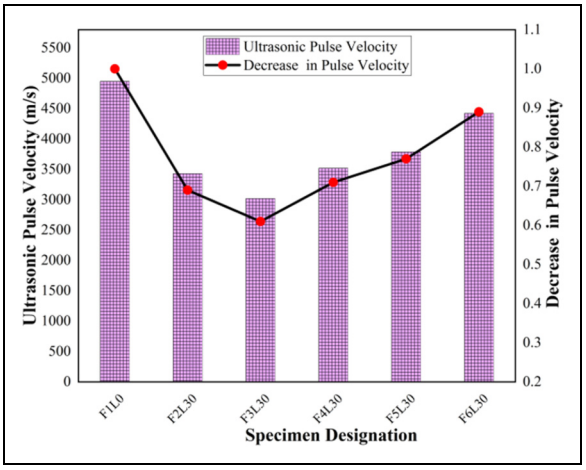


Figure 6. Dosage versus decrease in pulse velocity.

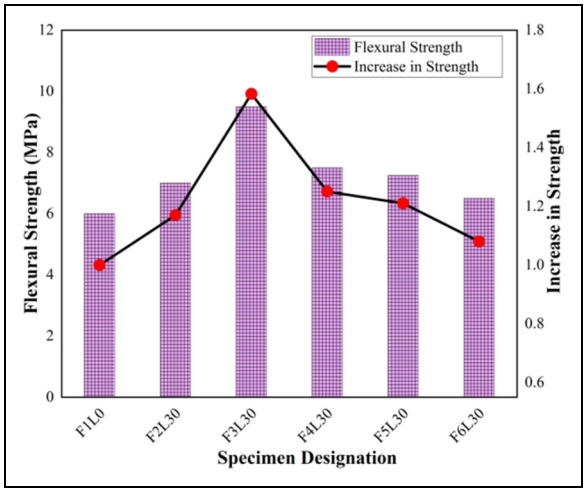


Figure 8. Dosage versus increase in strength.

increased with the fibre dosage increment after the point when there was a decline in pulse velocity values of samples until the optimum dosage point (F3L30) was attained. Still, the pulse velocity did not increase beyond this optimum dose point. This trend was found among the samples. The cast prismatic beam samples were subjected to a four-point bending test to determine their flexure strength. The results of this test revealed that the samples' flexural properties had been improved. When either the dosage or the Vf value was increased, as depicted in Figures 7 and 8, a significant increment was observed in the level of strength. Compared to the specimen that included no fibre in any percentage, the F3L30 sample showed an improvement in strength, around 58% higher than the other sample. Compared to F2L30, the reinforced samples of type F3L30 showed an improvement in strength of about 36% compared to those of type F2L30 (most negligible fibre dosage value). A pattern was observed, as indicated in Figures 7 and 8, that suggests that the flexural strength values increased up to an optimum dose point (0.4%) and then decreased later beyond this point. This pattern was observed to indicate that the value of the strength increases up to the limit. XRD analysis was the non-destructive method used on the reference sample (which had 0% fibre) and the F3L30 specimen, which included the optimal fibre dosage. This method was applied to the samples better to understand the mineralogical characteristics of the cast samples, as shown in Figure 9. It was noted that Ca–K and O–K were the significant peaks visible, which means that the samples used for testing are composed of silicon and calcium crystalline. This finding is more evidence that the primary products that result from the hydration of cement are calcium hydroxide and calcium silicate hydrate

XRD, SEM, and EDX

SEM was used to examine the surface morphology of the chopped carbon fibres and carbon fibre-reinforced concrete matrix used in the study, as illustrated in Figure 10. Figure 10(c) shows the surface of a chopped carbon fibre consisting of grooves that strengthen the interface binding between the fibres and the matrix. The mineral orientation and crystal microstructure of the cast concrete samples containing homogenously dispersed chopped carbon fibres

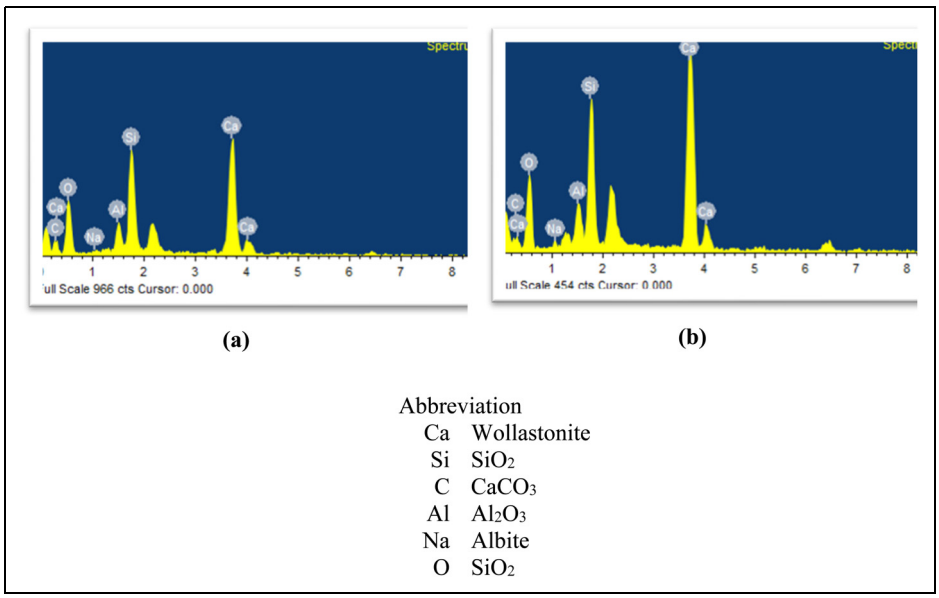


Figure 9. EDX results of samples (a) F1L0 and (b) F3L30.

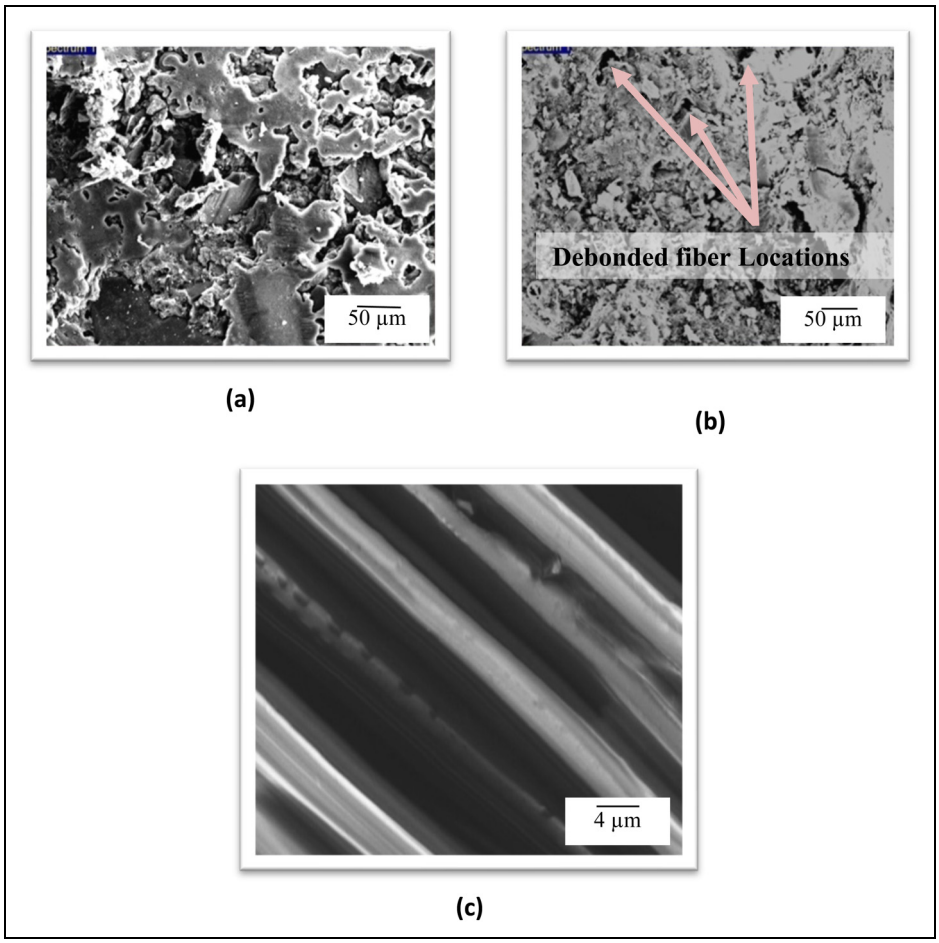


Figure 10. SEM images of (a) F1L0, (b) F3L30, and (c) View of carbon fibres magnified.

were observed in Figure 10(b) and (c). Debonding of carbon fibres from the concrete matrix was detected in Figure 10(b), implying that the carbon fibres employed in the experiments could not withstand the ultimate loads. Energy dispersive X-ray spectroscopy (EDX), as presented in Figure 9, described the abundance of elements present in the cast samples. Various elements found in the cast samples are listed in Table 3. As can be seen in

Table 3. Compositional content of prepared concrete samples in terms of percentages for several elements.

	F1L0		F3L30	
Element	Weight (%)	Atomic (%)	Weight (%)	Atomic (%)
O–K	49.38	58.71	49.24	62.43
C–K	12.48	19.77	6.76	11.42
Na–K	0.31	0.26	0.77	0.68
Al–K	3.55	2.50	3.64	2.74
Si–K	12.27	8.31	12.49	9.02
Ca–K	22.01	10.45	27.10	13.72
Total	100		100	

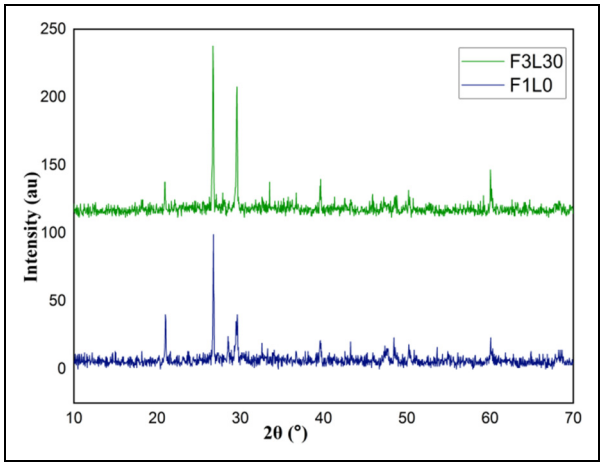


Figure 11. X-ray diffraction (XRD) analysis of concrete samples F1L0 and F3L30.

Figure 11, XRD analysis was done on concrete samples F1L0 and F3L30 to understand their mineralogical properties better. The XRD analysis of the concrete matrix revealed that oxygen (O) and calcium (Ca) elements were present in dominance in the matrix silicon, followed by comparable amounts of silicon (Si) and carbon (C) and only minute quantities of aluminium and sodium. Based on the EDX analysis, it was implied that calcium silicate hydrate and calcium hydroxide were the primary hydration products of the cement-based matrix.

Analytical investigation

The results of the four-point bending test were used to determine the flexural strength values of each of the particular specimens, which was done as part of the analytical study. The technique employed to achieve this investigation was linear regression analysis. The different composition Vf values of the cast prism samples were linearly correlated with the obtained Fr values. As depicted in Figure 12, an empirical link was constructed between these two parameters, denoted by the acronyms Vf and Fr. Here, Vf is the dosage value or the volume fraction of carbon fibres used, and Fr is the flexural strength values obtained from the flexure test. A marginal coefficient of determination value was obtained from the empirical relation.

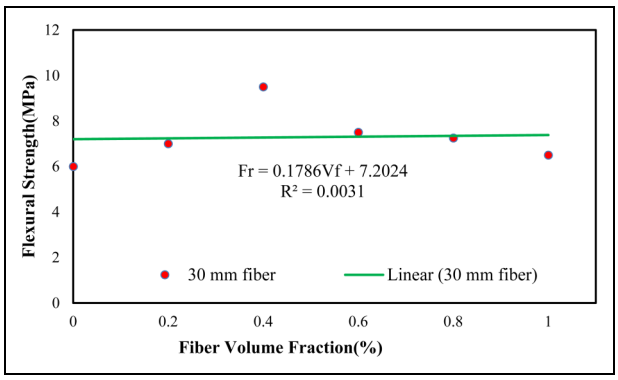


Figure 12. Correlation of Fr with volume fraction (Vf).

Conclusions

To understand the flexure performance of high-strength concrete, a sufficient amount of research was conducted on samples prepared by reinforcing chopped fibres of different mixes with different doses in concrete. This research was carried out to understand the flexure performance of high-strength concrete. A non-destructive test was performed using ultrasonic pulse velocity test equipment, a flexure test on cast specimens, an XRD study, and an examination of the material's composition. The following important points can be deduced from this inquiry and their findings:

- As the values of doses were increased, there was an increase in flexural performance; however, this gain was only observed up to an optimum dose point of fibre-reinforced concrete; after reaching this value, the strength values began to decline.
- A significant drop in pulse velocity was noticed up to an optimum dosage value, and pulse velocity values increased beyond this point.
- A trend was observed that the flexural strength values first increased and then decreased later, and this trend was reversed with ultrasonic pulse velocity test values.
- From the XRD result, it was clear that the peaks obtained were of oxygen (O) and calcium (Ca) elements, which were predominant in the matrix silicon, followed by comparable amounts of silicon (Si) and carbon (C), and only minute quantities of aluminium and sodium
- Debonding was shown to be the most common cause of failure in chopped carbon fibres by scanning electron microscope images.
- There was a marginal value of the coefficient of determination value when flexure strength values were correlated with different dosage values.

Acknowledgements

The authors would like to thank the support staff of civil engineering and applied physics research laboratories of DTU, Delhi.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

1. Ganesh M and Hariharan AR. Study on strength development of high strength concrete containing fly ash and silica fume. *Int J Eng Sci Technol* 2011; 9: 3619–3625.
2. Dong S, Zhou D, Ashour A, et al. Flexural toughness and calculation model of super-fine stainless wire reinforced reactive powder concrete. *Cem Concr Compos* 2019; 104: 103367.
3. Han B, Zhang L, Zhang C, et al. Reinforcement effect and mechanism of carbon fibers to mechanical and electrically conductive properties of cement-based materials. *Constr Build Mater* 2016; 125: 479–489.
4. Han B, Dong S, Ou J, et al. Microstructure related mechanical behaviors of short-cut super-fine stainless wire reinforced reactive powder concrete. *Mater Des* 2016; 96: 16–26.
5. Ghernouti Y, Rabehi B, Bouziani T, et al. Fresh and hardened properties of self-compacting concrete containing plastic bag waste fibers (WFSCC). *Constr Build Mater* 2015; 82: 89–100.
6. Mastali M and Dalvand A. The impact resistance and mechanical properties of self-compacting concrete reinforced with recycled CFRP pieces. *Compos B Eng* 2016; 92: 360–376.
7. Khaloo AR, Esrafili A, Kalani M, et al. Use of polymer fibres recovered from waste car timing belts in high performance concrete. *Constr Build Mater* 2015; 80: 31–37.
8. Mastali M, Dalvand A and Sattarifard AR. The impact resistance and mechanical properties of reinforced self-compacting concrete with recycled glass fibre reinforced polymers. *J Clean Prod* 2016; 124: 312–324.
9. Mastali M and Dalvand A. Use of silica fume and recycled steel fibers in self-compacting concrete (SCC). *Constr Build Mater* 2016; 125: 196–209.
10. Meddah MS and Bencheikh M. Properties of concrete reinforced with different kinds of industrial waste fibre materials. *Constr Build Mater* 2009; 23: 3196–3205.
11. Yang S, Yue X, Liu X, et al. Properties of self-compacting lightweight concrete containing recycled plastic particles. *Constr Build Mater* 2015; 84: 444–453.
12. Ochi T, Okubo S and Fukui K. Development of recycled PET fiber and its application as concrete-reinforcing fiber. *Cem Concr Compos* 2007; 29: 448–455.
13. Ogi K, Shinoda T and Mizui M. Strength in concrete reinforced with recycled CFRP pieces. *Compos A: Appl Sci Manuf* 2005; 36: 893–902.
14. Foti D. Use of recycled waste pet bottles fibers for the reinforcement of concrete. *Compos Struct* 2013; 96: 396–404.
15. Foti D. Preliminary analysis of concrete reinforced with waste bottles PET fibers. *Constr Build Mater* 2011; 25: 1906–1915.
16. Kim SB, Yi NH, Kim HY, et al. Material and structural performance evaluation of recycled PET fiber reinforced concrete. *Cem Concr Compos* 2010; 32: 232–240.
17. Lamba N, Raj R and Singh P. Mechanical characteristics of high strength concrete incorporating recycled CFRP fibers. *J Appl Polym Sci* 2022; 139: e53183.
18. Lamba N, Raj R and Singh P. Mechanical response of recycled carbon fiber reinforced polymer fibers in high-strength concrete. *Mater Today Proc* 2022; 78: 603–607.
19. IS: 8112-2013. Specification for 43 grade ordinary Portland cement. New Delhi: Bureau of Indian Standards, 2013, 17 p.
20. IS: 3812 1981. Indian standard specification for fly ash for use as pozzolana and admixture. New Delhi: Bureau of Indian Standards, 1999; 1 p.
21. ASTM C1240. Standard specification for silica fume used in cementitious mixtures. Annual Book of ASTM Standards 2020; 1–7.
22. AASHTO Designation: M 307-13. *Standard specification for silica fume used in cementitious mixtures*. Washington, DC: American Association of State Highway and Transportation Officials, 2021; vol. 13, pp. 1–7.
23. IS:15388. Silica fume-specification. New Delhi, India: Bureau of Indian Standards, 2003; pp. 1–13.
24. IS 9103. Specification for concrete admixtures. New Delhi, India: Bureau of Indian Standards, 1999; pp. 1–22.
25. Shetty MS. *Concrete technology theory and practice*. Multicolour Illustrative Edition. New Delhi: S Chand & Company Ltd, 2000; 55 pp.
26. IS:383. Specification for coarse and fine aggregates from natural sources for concrete. New Delhi: Bureau of Indian Standards, 1970, pp. 1–24.
27. IS: 2386 (Part I). Method of test for aggregate for concrete. Part I – Particle size and shape. New Delhi: Bureau of Indian Standards, 1963 (Reaffirmed 2002).
28. IS: 2386 (Part IV). Methods of test for aggregates for concrete, part 4: Mechanical properties. New Delhi: Bureau of Indian Standards, 2016, pp. 1–37.
29. IS: 2386 (Part III). Method of test for aggregate for concrete. Part III – specific gravity, density, voids, absorption and bulking. New Delhi: Bureau of Indian Standards, 1963 (Reaffirmed 2002).
30. IS: 13311 (Part 1). Method of non-destructive testing of concrete, part 1: Ultrasonic pulse velocity. New Delhi: Bureau of Indian Standards, 1992, pp. 1–7.
31. ASTM 2010. ASTM C78/C78 M. *Standard test method for flexural strength of concrete (using simple beam with third-point loading)*. West Conshohocken, PA: American Society for Testing and Materials, 2010, 100 pp.

Electrical Engineering
https://doi.org/10.1007/s00202-024-02239-5

ORIGINAL PAPER

Coal power plant-enabled grid resilience through distributed energy resources and demand response integration

Vivek Saxena[1,2] · Narendra Kumar[1] · Uma Nangia[1]

Received: 13 September 2023 / Accepted: 2 January 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
In the growing world, the utilization of electrical energy is increasing rapidly. Excessive use of fossil fuels will drain them and also invite hazardous pollution. Integrating renewable energy resources as distributed generators (DGs) can fulfill the rapidly increasing electrical energy demand and promote green energy generation to a large extent. The intermittent nature of renewable energy and higher penetration of DG may adversely affect the operation of the distribution network (DN). As a result, power disparity, reverse power flow to the grid, and voltage instability may exist. One key solution is to optimally integrate the renewable energy-based DG and battery energy storage system (BESS) in the coordination of demand response (DR). This paper proposes a multilevel particle swarm optimization technique to synchronize the distributed energy resources (DER) and DR in the DN. The proposed approach is implemented on the IEEE 33 bus system energized by coal power plant (CPP). The first level of optimization finds the sizes and locations of DER (DG and BESS), and the next level determines the optimal power dispatch in the coordination of DR. The outcomes of this framework exhibit effectiveness in the optimal utilization of renewable energy resources and the enhancement of power quality parameters in DN so that CO_2 emissions are reduced by 32.71% from CPP.

Keywords CO_2 · Coal power plant · Distributed generation · Demand response · Renewable energy · Battery energy storage system · Multilevel optimization

List of symbols

t.p.	Time period
P_i^T, P_J^T	Active power magnitude ith and jth node for the t.p. of T
r_{ij}	Branch resistance between ith and jth node
$P^T_{\mathrm{BESS}\left(\frac{C_i}{D_i}\right)}$	Power of BESS (charging and discharging) at ith node for t.p. of T
I^T_{Grid}	Output current the substation transformer at any time T
$P^{\mathrm{Max.}}_{\mathrm{BESS}}$	Maximum limit of power dispatch
Ψ	Conversion factors from daily to yearly
η_c	Charging efficiency of BESS
P_c^T	Converter loss for t.p. of T
I_{ij}^T	Level of current between i and j bus in hour T
η_d	Discharging BESS efficiency
$P_{\mathrm{PV}}^{\max}$	Maximum size of DGs
$E_{\mathrm{BESS},i}$	BESS energy at ith node
V_{Max}, V_{Min}	Limits of node voltages
η	Efficiency of the storage system
$P^{\mathrm{Min.}}_{\mathrm{BESS}}$	Minimum limit of power dispatch
V_i^T	Magnitude of the voltage at the ith node during the t.p. of T
V^T_{Grid}	Magnitude of grid voltage at any time T
$E^{\mathrm{Max}}_{\mathrm{BESS}}$	Maximum limit of energy
C	Contract load
$I_{ij}^{\max}$	Thermal limit (maximum) of line between bus i and j (A)
K_i^T	Allocated load factor for bus i at t.p. of T
$P_{\mathrm{el},i}^{\min}$, $P_{\mathrm{el},i}^{\max}$	Limits of responsive load (maximum and minimum)

✉ Vivek Saxena
vvksaxena1234@gmail.com

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

[2] Department of Electrical and Electronics Engineering, A.B.E.S. Engineering College, Ghaziabad, India

Published online: 29 January 2024

V_i^T, V_j^T	Node voltage of *i*th and *j*th node for the t.p. of T
$P_{\text{in},i}^T$	Instantaneous load (nonresponsive and responsive)
Q_{Di}^T	Reactive power demand at node *i*th at any time T
P_{Loss}^T	Power delivery loss for the t.p. of T
Q_i^T, Q_j^T	Reactive power magnitude *i*th and *j*th node for the t.p. of T
R_{PV}^r	Rated PV module radiations
$P_{\text{el},i}^T$	Responsive load at any time T
P_{Di}^T	Real power demand at node *i*th at any time T
E_i^{Total}	Total demand of energy of the day
$\text{SOC}^{\text{Min.}}$, $\text{SOC}^{\text{Max.}}$	SOC limits
P_{Gi}^T	Real power injection at node *i*th
R_{PV}^T	Solar radiation at time T
χ	Penetration of DR
SOC_i^T	SOC of BESS at bus i in hour T
$P_{\text{PV},i}$	Active power injection at node *i*th at any time T
V_D^T	Voltage deviation penalty
I_{Spc}	Specified reverse current limit
P_R^T	Reverse power flow at any time T
Δt	Change in time
I_{PV}	Solar module current (A)
δ_i^T, δ_j^T	Voltage angle of *i*th and *j*th node for the t.p. of T
Q_{Gi}^T	Reactive power injection at node *i*th for any time T
$L_{d,i}^T$	Hourly demand for the t.p. of T
$P_{D,i}^0$	Initial real power demand at *i*th node
$Q_{D,i}^0$	Initial reactive power demand at *i*th node
N	Number of buses

1 Introduction

In the modern age, the integration of renewable energy resources demands the implementation of a smart grid, utilizing real-time information and advanced communication technologies to achieve its objectives. Smart grids enable bidirectional communication between energy consumers and producers [1]. Several studies have investigated responsive load management based on dynamic pricing, all arriving at the same conclusion: shifting energy consumption from peak to off-peak hours in coordination with renewable energy resources maximizes benefits for prosumers [2–4]. Rapid advancements in small-scale generation technologies have compelled distribution system operators (DSOs) to increase the proportion of distributed generators (DGs) in distribution networks (DN). To optimize DN performance, it is essential to determine the ideal size and location of DGs. Previous research has demonstrated that suboptimal resource allocation can have counterproductive effects [5–7].

By incorporating DGs optimally, various objectives have been achieved, such as enhancing power quality, improving voltage profile, reducing atmospheric pollutants, and enhancing system stability and reliability, among others. However, traditional distribution systems face challenges in handling a large number of renewable energy sources due to their uncontrollable and unpredictable nature, which also varies over time [8].

To address this, one potential solution is the adoption of battery energy storage systems (BESS) to increase the penetration of power from non-dispatchable DGs in the DN [9–12]. Many studies have focused on modeling and optimizing the energy management of BESS to support DG integration.

While considerable efforts have been made to optimize BESS capacity, the location aspect is often overlooked in the objective function [9, 10]. To fully optimize DN performance, it is highly recommended to simultaneously consider both the location and size of BESS, which currently lacks adequate attention in the academic literature. By identifying the optimal size and position of BESS, the benefits of DSOs can be significantly enhanced [11].

The optimization of BESS involves various objectives, including minimizing power losses, depreciation of cost functions, and maximizing energy arbitrage profits, as discussed in [12]. The study evaluates the implications of optimizing the capacity and location of BESS for hybrid power plants based on wind energy and hydro energy in coordination with BESS, as stated in [13].

In their research [14], the authors explored how network reconfiguration influences the sizing and placement of BESS. However, implementing BESS can be costly for the DSO, leading to increased investment and operational expenses. The lifespan of BESS is influenced by factors such as the depth of discharge (DOD) and charging and discharging cycles, as mentioned in [15]. Studies indicate that a well-positioned and sizable BESS can support high penetration of DGs. Yet, expanding the BESS size can create additional cost burdens for the DSO, prompting researchers to explore alternative methods for DR to alleviate this pressure on BESS. DR has shown promise in various aspects of the smart grid, including communication and consumer coordination through price-based demand response, leading to reduced energy consumption costs and peak occurrence [16]. DR implementation in the DN in Finland resulted in technological benefits like reduced power losses and improved voltage profiles [17]. Researchers demonstrated a DR-coordinated

approach in allocating DGs to mitigate the intermittent nature of renewable energy sources [18–21].

The authors have effectively showcased the efficiency and advantages of integrating renewable energy planning into the coordinated management of electric vehicles [22, 23]. This integration not only contributes to the optimization of the DN but also emphasizes the intersection of technological advancements and economic benefits. To realize these synergies within the DN, the study employs sophisticated bilevel or multilevel optimization approaches [24–26]. These approaches, characterized by their hierarchical structure, enable a comprehensive exploration of technological and economic optimizations in hybrid energy systems [27–30].

Furthermore, the study recognizes the pivotal role of DER in enhancing the power grid's overall performance. The integration of DER into the power grid has been documented to significantly improve various power quality parameters [31–33]. This underscores the broader impact of incorporating DER technologies, such as renewable energy sources, in bolstering the reliability and quality of the electrical power supply.

DR is implemented to address demand and supply imbalances and mitigate the constraints faced by DG owners. In a previous study by the authors [34], they explored the joint deployment of DR and DG loads in the DN. The authors demonstrated a coordinated approach to optimize the benefits for various stakeholders in the utility sector [35, 36]. Based on this research, it was found that DR proves effective in scenarios with regulated loads, dynamic pricing, and renewable resources. After an extensive analysis of relevant published research, the authors concluded that integrating DGs and BESS into distribution networks significantly improve network efficiency. DR facilitates a high penetration of DGs and provides substantial advantages to smart grid participants, including customers, DSOs, generating firms, and aggregators. Although previous research has explored the optimal allocation of DGs when coordinating with DR, the incorporation of BESS in this context remains understudied. Hence, the objective of this research is to examine the role and benefits of DR in effectively integrating DGs and BESS in the DN. A key factor influencing the effectiveness of DR is consumer participation levels. Therefore, the study aims to determine how DR rates impact the appropriate sizing and placement of solar photo voltaic (SPV) and BESS systems. To achieve this, a multilevel optimization framework is utilized to seamlessly integrate PVs and BESS with DR planning in the DN.

According to the report released by the Central Electricity Authority (CEA) in 2022, the carbon dioxide (CO_2) emission from coal power plants (CPP) in the Indian power sector is 0.975 tCO_2/MWh [37]. To reduce these emissions, strategically placing renewable DG in the DN is essential. In this paper, SPV is considered as the renewable DG, known for its pollution-free electricity production. However, it is crucial to note that CO_2 emissions occur during the construction of SPV plant components, as they rely on electricity from thermal power plants. For SPV modules, the estimated lifetime CO_2 emission is 0.053 kg per 1 kWh of electrical energy [38].

Figure 1 shows the capacity additions from financial year (FY) 2000–2001 to FY 2021–2022. Notably, coal-based capacity witnessed substantial growth from FY 2000–2001 to FY 2015–2016, but there was a significant decline in capacity additions from FY 2016–2017 to FY 2021–2022. Hydro-based capacity additions also experienced a notable decrease starting from FY 2017–2018. No recorded additions were observed in other generation capacities during this period.

Figure 2 illustrates the trend of the weighted average emission factor from FY 2017–2018 to FY 2021–2022. During FY 2021–2022, there was a slight increase in the weighted average emission factor, primarily due to a rise in total electricity generation, with approximately a 2% increase in coal-based generation. However, there was a minor reduction in gas and hydro-based generation. Additionally, the proportion of imported coal decreased from 9 to 4% compared to the previous fiscal year.

This research paper encompasses several objectives, including minimizing feeder energy losses and BESS energy conversions, reducing voltage deviations, and minimizing reverse power flow while maximizing the use of BESS and DR in the presence of high SPV penetration. So that the CO_2 emission can be reduced effectively. To assess the effectiveness of proposed framework, numerous test scenarios have been examined to a typical 33-bus test data set and the power grid is supplied by CPP.

2 Problem conceptualization

Effectively increasing the adoption of renewable energy sources and optimizing energy efficiency can lead to substantial reductions in CO_2 emissions. By reducing the electricity demand from conventional CPP, significant CO_2 emissions can also be curtailed.

In this study, the following objectives have been examined for the optimal allocation of PV and BESS in the effective synchronization of DR panning in DN.

2.1 Minimization of power losses

Power losses during power transmission in DN are considerable. Implementing technological advancements like integrating DGs, optimizing BESS functionality, and coordinating DR offer substantial potential for reducing these losses. Minimizing power loss is a crucial goal function, as it

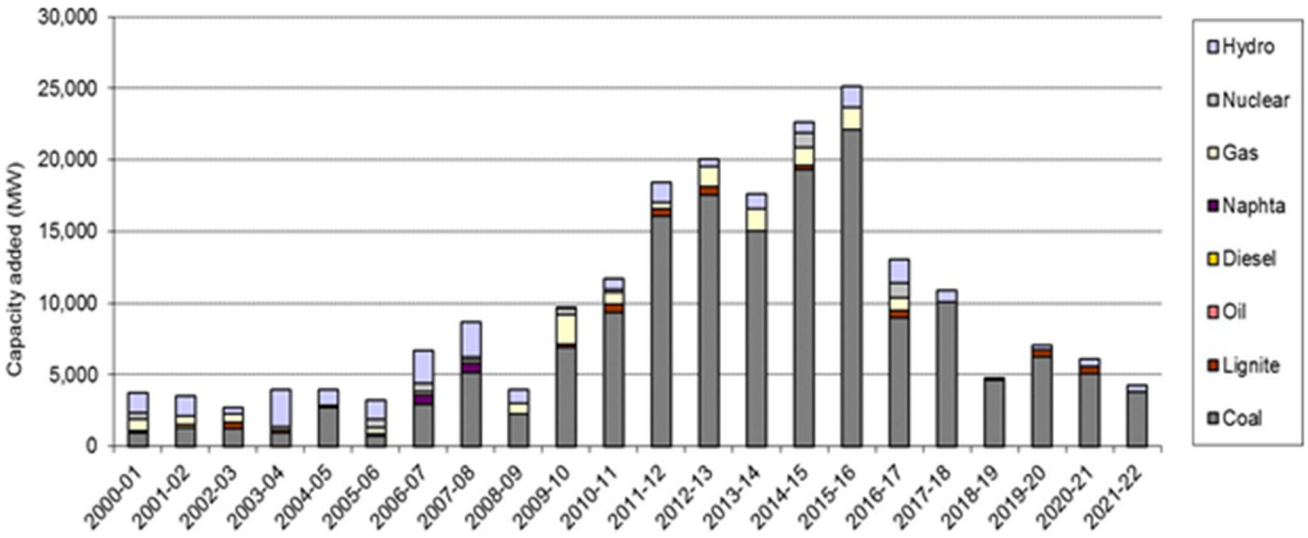


Fig. 1 Breakdown of new added capacity covered by the database over the period 2000–2001 to 2021–2022 [37]

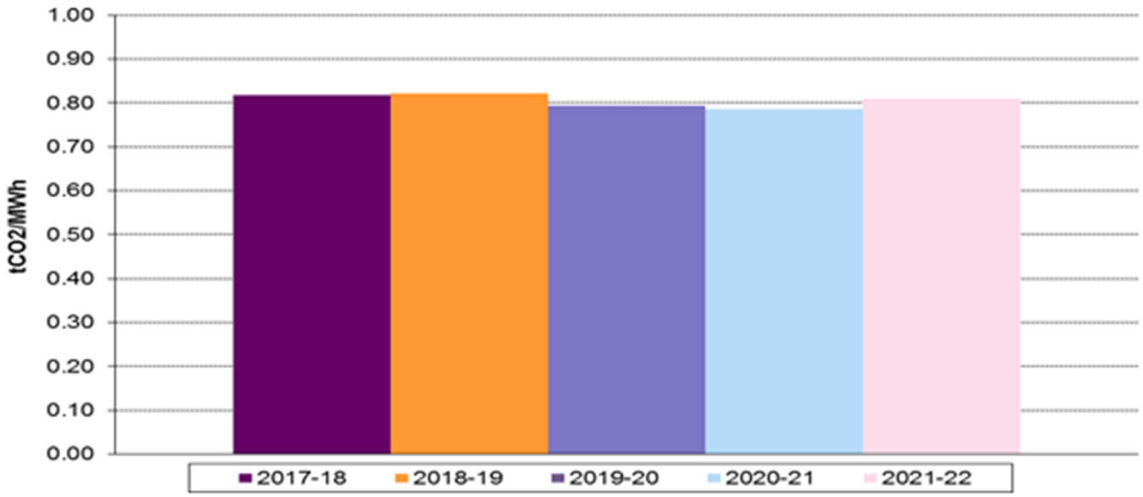


Fig. 2 Development of the weighted average emission factor for the Indian Grid over the period 2017–2018 to 2021–2022 [37]

directly affects the annual income of utilities and has implications for all stakeholders [39].

$$f_1 = \sum_{T=1}^{24} P_{\text{Loss}}^T \tag{1}$$

$$P_{\text{Loss}}^T = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_{ij}^T \left(P_i^T P_J^T + Q_i^T Q_J^T\right) + \beta_{ij}^T \left(Q_i^T P_j^T - P_i^T Q_j^T\right) \forall T \tag{2}$$

where $\alpha_{ij}^T = r_{ij}\cos\left(\delta_i^T - \delta_j^T\right)/V_i^T V_j^T$ and $\beta_{ij}^T = r_{ij}\sin\left(\delta_i^T - \delta_j^T\right)/V_i^T V_J^T$.

2.2 Minimization of reverse power flow

Anticipated rise in renewables to meet green energy targets and reduce emissions. Increasing penetration may cause reverse power flow during low demand, posing operational complexity and protection challenges for DSOs. DG integration aims to address this reverse flow possibility. The goal: achieving bidirectional power flow.

$$f_2 = \sum_{T=1}^{24} P_R^T \tag{3}$$

$$P_R^T = \begin{cases} 0, & \text{if } I_{\text{Grid}}^T \geq I_{\text{Spc}} \\ \text{Re}\left(V_{\text{Grid}}^T I_{\text{Grid}}^{T*}\right) & \text{if } I_{\text{Grid}}^T < I_{\text{Spc.}}. \end{cases} \tag{4}$$

2.3 Minimization of BESS conversion losses

Recent research has extensively focused on converters and battery storage systems. To ensure efficient integration of BESS, it is crucial to minimize converter losses. As a result, one of the objectives is to consider the conversion losses during the charging and discharging periods of BESS. The formula for BESS loss is as follows:

$$f_3 = \left(\sum_{T=1}^{24} P_c^T \right) \tag{5}$$

$$P_c^T = (1-\eta) P^T_{\mathrm{BESS}\left(\frac{C_i}{D_i}\right)} \tag{6}$$

2.4 Node voltage deviation

To maintain power supply reliability, the DN must operate within an acceptable voltage range. Typically, a penalty function is used to account for voltage violations. This penalty function can also be considered as one of the objectives.

The purpose of exceeding the voltage limit can be stated as follows [40]:

$$f_4 = \left(1 + \sum_{T=1}^{24} V_D^T \right) \tag{7}$$

$$V_D^T = \begin{cases} \left| V_{\mathrm{Min}} - V_i^T \right| & \text{if } V_i^T < V_{\mathrm{Min.}} \\ 0 & \text{if } V_{\mathrm{Min.}} \le V_i^T \le V_{\mathrm{Max.}} \\ \ell & \text{if } V_i^T > V_{\mathrm{Max.}} \end{cases} \tag{8}$$

$\boldsymbol{l}$ denotes an arbitrarily large positive integer. A high value represents an unfavorable solution, particularly when it comes to voltage rise beyond the maximum working limit, which poses a significant problem [41]. The issue of voltage rise intensifies during low demand, potentially leading to equipment failure due to insulation collapse and impact on protection measures. The increasing voltage poses a major obstacle to deeper DG penetration in the DN. To address this, a hard constraint is imposed to prevent violations of the higher voltage limit. Conversely, the minimum voltage is treated with more flexibility, allowing slight deviations as long as an appropriate penalty factor is applied.

2.5 Fitness function

To achieve the required objectives, it is required to make a fitness function having the weightage factor of different objective functions. For level 1 of optimization, the fitness function ($F_{\ell 1}$) is given as:

$$\min(F_{\ell 1}) = \Psi \times L \times f_4 \tag{9}$$

where $L = f_1 + f_2 + f_3$.

The level 2 of the optimization objectives involves taking into consideration the DR planning and scheduling methods of DGs and BESS. In this stage, the BESS conversion losses are irrelevant and have no impact. Because of this, the following objective function will be taken into consideration for level 2 of the optimization problem:

$$\min(F_{\ell 2}) = \kappa \times f_4 \tag{10}$$

where $\kappa = f_1 + f_2$.

In this context, the fitness function for level 2 is denoted by $F_{\ell 2}$.

It is vital to have a dispatch strategy, which is determined upon by the DR aggregator. This helps to minimize the aforementioned goal function.

2.6 Demand response aggregator

The developed DR system considers the benefits enjoyed by various stakeholders, including DSOs and consumers. The DR aggregator influences customers' demand scheduling based on their participation in DR, subject to technological limitations. This study primarily focuses on the technological challenges caused by the widespread use of SPVs. DR implementation uses dynamic tariff information to schedule loads and achieve maximum reduction in set technical targets, without considering financial objectives. Participants in mandatory DR programs may face financial penalties for not following the aggregator's instructions. The scheduling of demand aims to balance total consumption and available resources throughout the day, restructuring demand instead of simply reducing overall consumption.

The DR aggregator schedules load by collecting information on responsive and nonresponsive load demand, making it an essential part of the program.

Responsive load demand refers to the electricity demand that can be adjusted or shifted in response to instructions from the DSO. On the other hand, nonresponsive load demand denotes the electricity demand that cannot be altered or shifted from its usual operating hours. Essentially, responsive loads provide flexibility in their consumption patterns, allowing for adjustments to optimize the overall performance and efficiency of the electrical grid, particularly in scenarios where the DSO seeks to balance supply and demand or manage grid constraints. In contrast, nonresponsive loads maintain a fixed demand profile, typically adhering to their established operating schedules without the capacity for immediate adjustment based on external instructions. Equation (13) represents the complete demand at any time T, which

is the sum of all types of loads (responsive and nonresponsive). To meet responsive demand without affecting overall daily demand, scheduling restrictions are defined in Eq. (14), with Eq. (15) representing the minimum and maximum values of responsive demand. The peak value of responsive demand depends on the penetration level of DR, as shown in Eq. (16). Various DR restrictions are considered, as discussed in [42].

$$P_i^T = \left(P_{Gi}^T - P_{Di}^T\right)\forall T,\ i \tag{11}$$

$$Q_i^T = \left(Q_{Gi}^T - Q_{Di}^T\right)\forall T,\ \mathrm{I} \tag{12}$$

$$P_{Di}^T = \left(P_{\mathrm{in},i}^T + P_{\mathrm{el},i}^T\right)\forall T,\ i \tag{13}$$

$$\sum_{i=1}^{N}\sum_{T=1}^{24}\left(P_{\mathrm{in},i}^T + P_{\mathrm{el},i}^T\right) \times \Delta t = E_i^{\mathrm{Total}} \tag{14}$$

$$P_{\mathrm{el},i}^{\min} \le P_{\mathrm{el},i}^T \le \min\left(\left(C - P_{\mathrm{in},i}^T\right),\ P_{\mathrm{el},i}^{\max}\right)\forall T \tag{15}$$

$$P_{\mathrm{el},i}^{\max} = \chi \sum_{T=1}^{24} L_{d,i}^T \tag{16}$$

The optimal solution to the optimization issue is a vector of responsive load schedules with a length of T and is represented as:

$$P_{\mathrm{el}} = \left[P_{\mathrm{el}}^1,\ P_{\mathrm{el}}^2,\ P_{\mathrm{el}}^3,\ P_{\mathrm{el}}^4 \cdots P_{\mathrm{el}}^T\right]\forall T.$$

2.7 Objective constraints

The objective functions are constrained in a variety of ways by both technical and operational considerations. These restrictions can be represented numerically as follows:

2.7.1 SPV generation limit constraint

The constraint for PV generation limit is given as:

$$0 \le P_{\mathrm{PV},i} \le P_{PV}^{\max}\forall i \tag{17}$$

2.7.2 BESS constraints

The constraints of BESS are given as:

$$0 \le E_{\mathrm{BESS},i} \le E_{\mathrm{BESS}}^{\mathrm{Max}}\forall i \tag{18}$$

$$P_{\mathrm{BESS}}^{\mathrm{Min.}} \le P_{\mathrm{BESS}(C_i/D_i)}^T \le P_{\mathrm{BESS}}^{\mathrm{Max.}}\forall T,\ i \tag{19}$$

$$\mathrm{SOC}^{\mathrm{Min.}} \le \mathrm{SOC}_i^T \le \mathrm{SOC}^{\mathrm{Max.}}\ \forall T,\ \mathrm{I} \tag{20}$$

$$\mathrm{SOC}_i^T = \begin{cases} \mathrm{SOC}_i^{T-1} + P_{\mathrm{BESS}(\mathrm{C}_i/\mathrm{D}_i)}^T \eta_c \Delta t / E_{\mathrm{BESS}}^R & \text{if } P_{\mathrm{BESS}(\mathrm{C}_i/\mathrm{D}_i)}^T > 0 \\ \mathrm{SOC}_i^{T-1} + P_{\mathrm{BESS}(\mathrm{C}_i/\mathrm{D}_i)}^T \Delta t / \eta_d E_{\mathrm{BESS}}^R & \text{else} \end{cases} \tag{21}$$

$$\sum_{T=1}^{24} \eta_c P_{\mathrm{BESS}(C_i/D_i)}^T + P_{\mathrm{BESS}(C_i/D_i)}^T / \eta_d = 0 \tag{22}$$

Equation (18) represents the limitations of energy while Eq. (19) represents the limitations of power dispatch. The limits of SOC are given in Eq. (20), and SOC status is presented in Eq. (21). SOC balancing constraints are demonstrated in Eq. (22). All the above equations are at specific node and time.

2.7.3 Feeder constraint

The constraint for the thermal limits is given as:

$$I_{ij}^T \le I_{ij}^{\max}\forall T,\ i,\ j \tag{23}$$

2.7.4 Power balance constraints

$$P_i^T = V_i^T \sum_{j=1}^{n} V_j^T Y_{ij} \cos\left(\theta_{ij} + \delta_j^T - \delta_i^T\right)\forall T,\ i \tag{24}$$

$$Q_i^T = -V_i^T \sum_{j=1}^{N} V_j^T Y_{ij} \sin\left(\theta_{ij} + \delta_j^T - \delta_i^T\right)\forall T,\ i \tag{25}$$

The actual and reactive power balance restrictions are shown by Eqs. (24) and (25), respectively.

2.8 Demand modeling

The demand modeling of the system is given in the following equations:

$$P_{D,i}^T = \kappa_i^T P_{D,i}^0 \forall T,\ i \tag{26}$$

$$Q_{D,i}^T = \kappa_i^T Q_{D,i}^0 \forall T,\ i \tag{27}$$

2.9 SPV generation modeling

Solar power generation is dependent on a number of other factors as well, such as the type of panel and its area, the angle at which it is tilted, and the amount of solar radiation that is received. For the purposes of this study, during a specific

period of time, it is assumed that all other parameters remain unchanged. The transformation of the current in relation to the rated voltage may be found as follows:

$$I_{\mathrm{pv}}^{T} = \begin{cases} I_{\mathrm{PV}} & \text{if } R_{\mathrm{PV}}^{T} \geq R_{\mathrm{PV}}^{r} \\ I_{\mathrm{PV}} \times R_{\mathrm{PV}}^{T}/R_{\mathrm{PV}}^{r} & \text{if } R_{\mathrm{PV}}^{T} < R_{\mathrm{PV}}^{r} \end{cases} \tag{28}$$

3 Optimization technique

As discussed in previous case, the presented optimization objectives required such a optimization technique that can solve the complex nonlinear problem. The adoption of a multilevel optimization context is required by the presence of BESS since it takes into consideration both the limits associated with SOC levels and their accessibility. The optimal allocations of PVs and the BESS are determined at the first level of optimization. At the second level of optimization, the hourly power scheduling of BESS in the synchronization of DR programs is determined. This is done to make sure that the operational gains from DSO are used to their fullest extent.

Any evolutionary method can be utilized to address the difficult multilevel optimization issue that has to be solved. According to a review of the relevant published material, it has been determined that PSO is the method that is utilized most frequently for tackling the DG planning optimization problem [43]. As a result, the optimization goal presented in this work has been met at both levels by using PSO.

The PSO is an optimization approach that has the capability to search for a global or near-global solution to difficult optimization issues involving power systems, and it is a particle-based meta-heuristic technique. Here are the things that need to be done to improve the multilevel optimization that is being thought about:

I. Set the initial values for the parameters and variables that are used in level 1 optimization. It includes locations, sizes of PVs and BESS, and parameters of the proposed optimization technique.
II. Upgrade the sizes and locations of PV that have been determined heuristically.
III. Apply the calculated load factor κ_i^T to the P_{Di}^T and Q_{Di}^T for a period of 24 h.
IV. The level 2 gets the most up-to-date location, size of BESS P_{Gi}^T, P_{Di}^T, and Q_{Di}^T so that BESS and DR can be managed as well as possible.
V. For the level 2, the arrangement of responsive load and BESS power dispatch over a period of twenty-four hours is seen as the variables.
VI. At the level 2, the scheduling of BESS and DR is started subsequently getting the outcomes from first level subjected to various constraints of the system.
VII. Execute the load flow to find out how much power is lost and how much voltage is at each node.
VIII. Perform another round of updates to P_{Gi}^T (BESS discharging) and P_{Di}^T (BESS charging and also depends on the planning of responsive loads) in accordance with the BESS and DR schedules that were optimized in level 2.
IX. The level 1 controller receives upgraded variables like P_{Gi}^T, P_{Di}^T, SOC, and BESS power in order to optimize the objective of Eq. (9).
X. Perform the load flow to find out the energy losses in addition the level of voltages at different nodes, it is necessary to evaluate the performance of level-1 optimization function.
XI. Preserve the level 1 population with the highest fitness and its matching best population.

The flow-chart that is given in Fig. 3 illustrates both the upper and lower levels' structures in great detail.

The scheduling of BESS will depend upon the optimized value of dispatch power and the present value of SOC. The value of DOD is assumed to be 20% of the peak value of DOD and, primarily, the value of SOC is equal to the value of DOD. The optimal power dispatch and SOC will determine how the BESS charges and drains.

4 Results and discussion

The proposed multilevel optimization technique is implemented on the IEEE 33 bus system as shown in Fig. 4 [44]. In this research, the effects of DR technologies are shown and analyzed in order to solve the optimal power dispatch problem in a number of different scenarios and with a number of different constraints. The optimization objectives are solved with the help of proposed optimization techniques by using MATLAB software on i3 core processor having 12 GB RAM. PV is considered as DG during the optimal planning of power dispatch with the coordination of BESS and different DR rates. The details about the simulation settings for the multilevel optimization are given in Table 1. Additionally, the values of base voltage, nominal active demand, nominal reactive demand, power loss, $V_{\min}$, $V_{\max}$, $P_{\mathrm{BESS}}^{\mathrm{Min}}$, $P_{\mathrm{BESS}}^{\mathrm{Max}}$, $\mathrm{SOC}^{\mathrm{Min}}$, $\mathrm{SOC}^{\mathrm{Max.}}$, $\mathrm{DG}^{\mathrm{Max.}}$ and $E_{\mathrm{BESS}}^{\mathrm{Max}}$ are 12.66 kV, 3715 kW, 2300 kVAr, 202.7 kW, 0.95 pu., 1.05 pu, − 1 MW, 1 MW, 0,1, 2 MW and 5 MWh, respectively.

Table 2 demonstrates how the DR rates impact the overall performance of the DN. The participation of different customers in various DR programs is reflected in the various DR

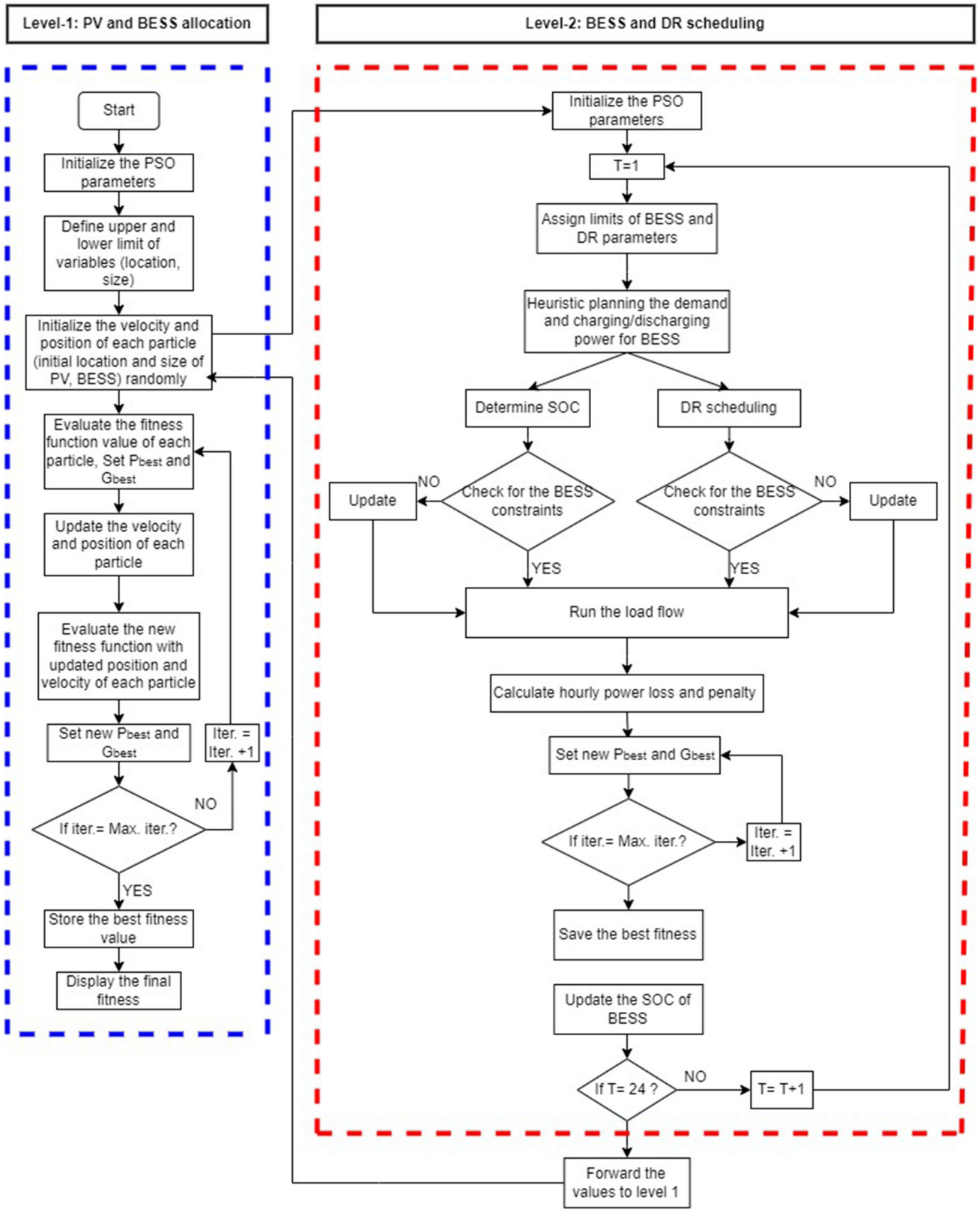


Fig. 3 Flow-chart for multilevel optimization approach

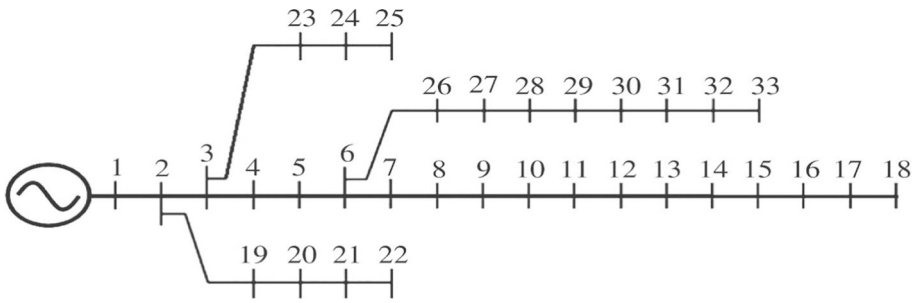


Fig. 4 IEEE 33 bus system

Table 1 Simulation parameters of multilevel optimization technique

Parameters	Level-1	Level-2
Swarm size	20	50
Inertia weight	1	1
Inertia weight damping ratio	0.99	0.99
Personal learning coefficient	1.5	1.5
Global learning coefficient	2	2
Maximum number of iterations	50	50

rates. In this case study, we are assuming that type DR programs are required. Participants in mandated DR programs are required to pay the fine/penalty if they did not coordinate their usage in accordance with the DR aggregator's instructions.

As a result, it is essential for the DR aggregator to monitor, control, and advise the consumers on scheduling the demand in accordance with the implemented DR rate. The outcomes of the optimal planning of power dispatch in the coordination of DG, DR, and BESS are given in Table 3. Once the recommended technique is put into place, significant reductions in yearly energy loss, reverse power flow, and voltage variations are seen. As an illustration of the effectiveness of the suggested technique, we present several case studies and their respective results. The proposed approach has a noteworthy impact on the CO_2 emissions of the CPP, as clearly outlined in Table 4.

4.1 Case 1

The purpose of discussing this scenario is to showcase the effectiveness of the proposed approach. For this analysis, neither PV nor BESS is considered in the 33-bus radial DS. In this specific case, the DSO relies entirely on power from the grid. The daily consumption pattern, obtained from reference [45], is incorporated into the proposed objective functions for convenience and speed. Based on Table 2 and 3, the base scenario shows differences of 5397.73 kW between maximum and minimum demand, a minimum mean voltage of 0.978178 p.u., and yearly energy losses of 1426 MWh. The peak demand occurs around 20:00 in the evening, while the valley point is approximately at 5:00 in the morning. The total CO_2 emission per day is 75,446.33 kg.

Table 2 Effect of the coordination of DR with optimally integrated renewable DG and BESS on demand

Case no	Category	Maximum demand (kW)	Maximum demand mitigation %	Difference between maximum-to-minimum demand (kW)	% of maximum loss mitigation at 8:00 PM
1	Base Case	6519	0	5397.73	0
2	DG	6519	0	6016.39	0
3	DR@10%	5548	16.1	4166.14	36.09
	DR@20%	5321	18.42	3730.6	42.77
4	DG + DR@10%	5370	17.33	4322.87	33.69
	DG + DR@20%	4790	26.78	3540.31	45.69
5	DG + BESS	4355	33.23	4100.89	37.09
6	DG + BESS + DR@10%	3810	41.54	3299.7	49.38
	DG + BESS + DR@20%	3290	49.54	2498.53	61.67

Table 3 Outcomes of the coordination of DR with optimally integrated renewable DG and BESS

Case no	Category	Optimal allocation of DG (Bus No., kW)	Optimal allocation of BESS (Bus No., kWh)	Annual losses (MWh)	Reduced losses/year (%)	DG penetration (%)	Average voltage level (p.u.)
1	Base Case			1426			0.9781
2	DG	17(1344)-32(1690)-25(1092)		1098	23	68.76	0.9963
3	DR@10%			1302	8.69		0.9784
	DR@20%			1290	9.53		0.9785
4	DG + DR@10%	7(1086)-15(1902)-32(914)		996	30.15	65.03	0.9964
	DG + DR@20%	18(408)-29(1816)-11(1602)		934	34.5	63.76	0.9967
5	DG + BESS	18(1901)-30(1657)-7(1840)	13(2697)-33(4593)-16(3773)	810	43.19	89.96	1.0121
6	DG + BESS + DR@10%	16(1532)-32(1736)-26(1855)	16(4529)-33(4386)-18(982)	786	44.88	85.38	1.0123
	DG + BESS + DR@20%	17(998)-11(2024)-30(1747)	16(4591)-33(4388)-18(985)	724	49.22	79.48	1.0127

Table 4 Impact of proposed framework on CO_2 emission of CPP

Case	Energy demand from CPP/Day (kWh)	Energy supplied from DG/Day (kWh)	CO_2 emission from SPV (Kg)	Energy losses/day (kWh)	Energy supplied from CPP/Day (kWh)	CO_2 emission/Day (Kg) by CPP	Total CO_2 emission/Day (Kg)	% reduction in CO_2 emission/Day
Base Case	73,474			3906.85	77,380.85	75,446.33	75,446.33	
DG	52,681	20,793	1102.029	3008.22	55,689.22	54,296.99	55,399.02	26.51%
DR@10%	73,472			3567.12	77,039.12	75,113.15	75,113.15	0.30%
DR@20%	73,411			3534.25	76,945.25	75,021.62	75,021.62	0.49%
DG + DR@10%	50,684	22,790	1207.87	2728.77	53,412.77	52,077.45	53,285.32	29.31%
DG + DR@20%	54,498	18,976	1005.728	2558.90	57,056.90	55,630.48	56,636.21	24.87%
DG + BESS	48,640	24,834	1316.202	2265.00	50,905.00	49,632.38	50,948.58	32.47%
DG + BESS + DR@10%	48,640	24,834	1316.202	2161.60	50,801.60	49,531.56	50,847.76	32.60%
DG + BESS + DR@20%	48,640	24,834	1316.202	2079.45	50,719.45	49,451.46	50,767.67	32.71%

4.2 Case 2

In this case, the initial optimization level calculates the penetration level and locations of DGs while keeping the remaining information unchanged. According to Table 2 and 3, specific values for this scenario indicate the difference between maximum and minimum demand as 6016.39 kW, the minimum mean voltage as 0.99634 p.u., and the yearly energy losses as 1115 MWh. The peak demand remains constant, depicted in Fig. 2, owing to the difference in solar generation availability compared to peak demand occurrence. Another valley point observed from 10:00 to 15:00 is attributed to the increased penetration of DG, reducing grid demand.

The optimized integration of DGs significantly improves power quality parameters. Specifically, in this instance, yearly energy losses reduced by approximately 21.8%, and the minimum mean voltage increased from 0.978178 to 0.99634 p.u. Table 3 presents the optimal dimensions and placement of PV installations within DS. The impact of DGs on demand pattern, voltage pattern, and active power losses is illustrated in Figs. 5, 6, and 7, respectively. The CO_2 emission is reduced by 26.51% in comparison with base case scenario.

4.3 Case 3

In this specific scenario, the study focuses solely on evaluating the effectiveness of the DR approach without coordinating with DGs or BESS. Two different levels of demand elasticity, namely 10 and 20%, are considered as the DR rates. DR identifies the elastic load that can be adjusted to suit demand and pricing conditions, while the inelastic load remains constant. Introducing various elastic loads in DR significantly improve performance compared to the basic scenario. Peak demand is reduced by 14.72 and 18.32% for the 10 and 20% DR rates, respectively, and annual energy losses decrease between 5.96 and 8.2%. This confirms that DR leads to reduced energy losses and improved peak-to-valley difference, even without DG integration.

Figures 8 and 9 demonstrate the impact of a 10% DR rate on demand pattern, voltage pattern, and active power losses, respectively. Similarly, Figs. 10 and 11 show the impact of a 20% DR rate on the same aspects. Notably, in the absence of DG integration, DR has a negligible effect on the voltage profile and CO2 emission in comparison with the base case.

4.4 Case 4

In this scenario, the analysis focuses on incorporating DGs into DR coordination and planning, considering various system constraints. The integration of DGs with DR scheduling significantly enhances the benefits provided by DGs. Additionally, system performance improves with higher DR rates, even in the presence of smaller DGs. Notably, the yearly energy loss shows a substantial decrease, ranging from 29.03 to 33.31%, depending on the DR rates, and the lowest mean voltage experiences a significant increase compared to cases 1 and 2.

Furthermore, when comparing the results of using DGs alone, the difference between maximum-to-minimum demand decreases considerably, leading to a more balanced load profile. Figures 12, 13, and 14 demonstrate the impact of DGs and a 10% DR rate on demand, voltage, and active power losses, respectively. Similarly, Figs. 15, 16, and 17 illustrate the impact of DGs and a 20% DR rate on demand, voltage, and active power losses, respectively. The CO_2 emission for DR rate 10 and 20% are 29.31 and 24.87% respectively.

4.5 Case 5

A study was conducted to evaluate the impact of BESS on the optimal allocation of DGs. The findings, presented in Table 3, show the determined placement and dimensions of the BESS. The planning and scheduling of BESS charging and discharging were strategically executed to improve the voltage profile and reduce energy losses in the network. Coordinating the ideal BESS with DGs aimed to minimize the adverse effects of excessive penetration. The results in Table 2 demonstrate a significant reduction in yearly energy loss (down to 42%) and a flatter load curve compared to a previous case (case-2).

Furthermore, the application of BESS resulted in a remarkable increase in the optimum DG penetration (90.93% as opposed to 69.44%). Figures 18, 19, 20 and 21 showcase the impact of DGs and BESS on the demand pattern, voltage profile, active power losses, and BESS energy storage. The CO_2 emission is reduced by 32.47% in comparison with the base case.

4.6 Case 6

In this situation, the coordination between the DR aggregator and the BESS integrator is essential for optimizing the allocation of DGs and BESS. It is observed that the sizing of BESS and DG depends on the DR rate. Furthermore, a higher DR rate leads to smaller BESS and DG sizes, enhancing the efficiency of the system and improving the DN functionality. According to Table 3, the most significant reduction in yearly energy loss (46.77% savings) is achieved when combining DG, BESS, and a DR rate of 20%. This study emphasizes the importance of incorporating an appropriate DR rate alongside DGs and BESS to ensure an efficient operation of the DN. Figures 22, 23, 24, 25, 26, 27, 28 and 29 demonstrate the impact of varying DR rates on the synchronization of optimal DG and BESS allocation, including the effects on demand pattern, voltage profile, active

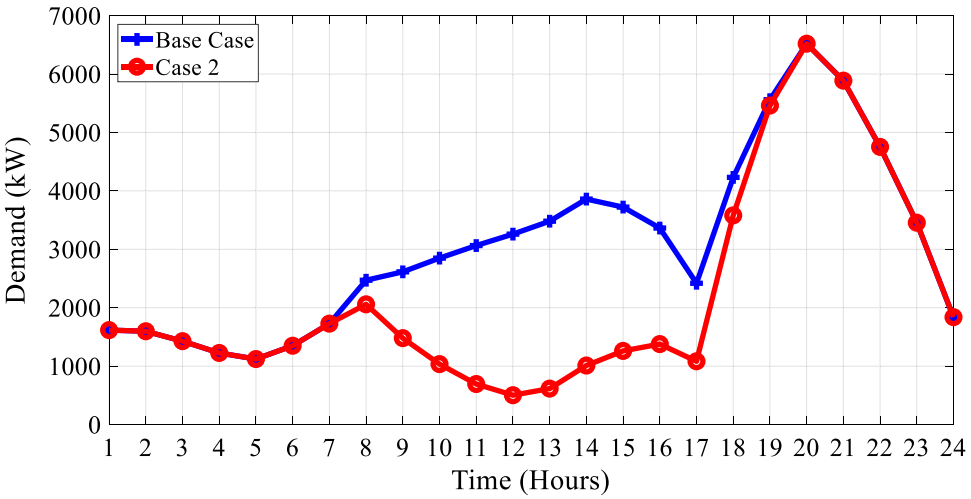


Fig. 5 Impact of DGs on demand pattern

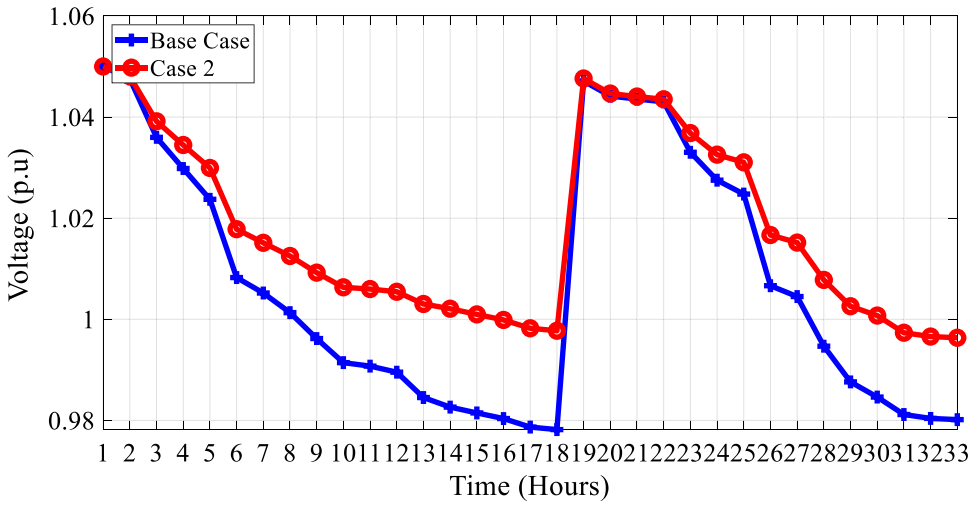


Fig. 6 Impact of DGs on voltage pattern

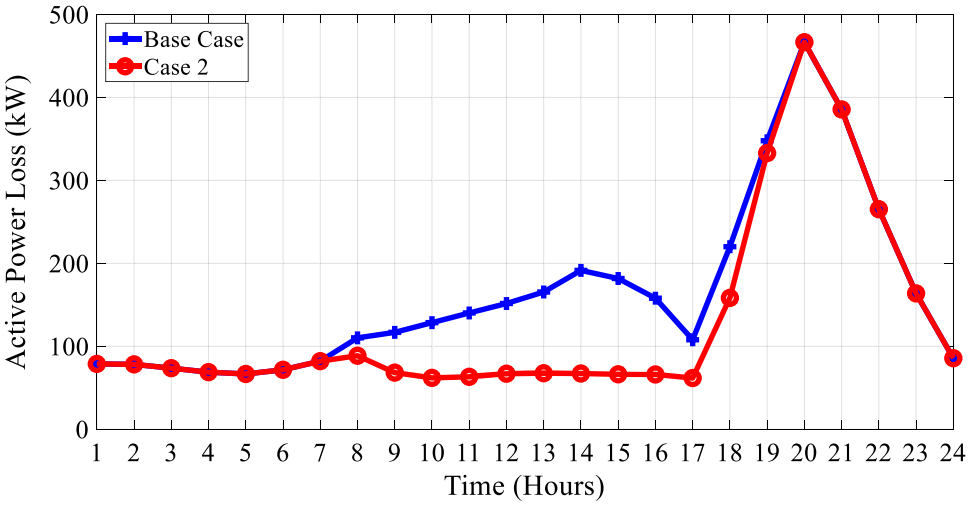


Fig. 7 Impact of DGs on active power losses

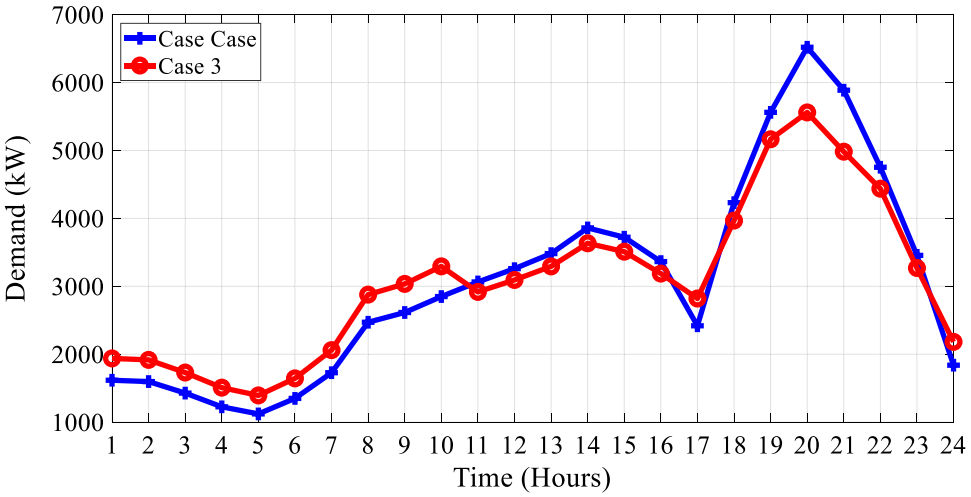


Fig. 8 Impact of 10% DR rate on demand pattern

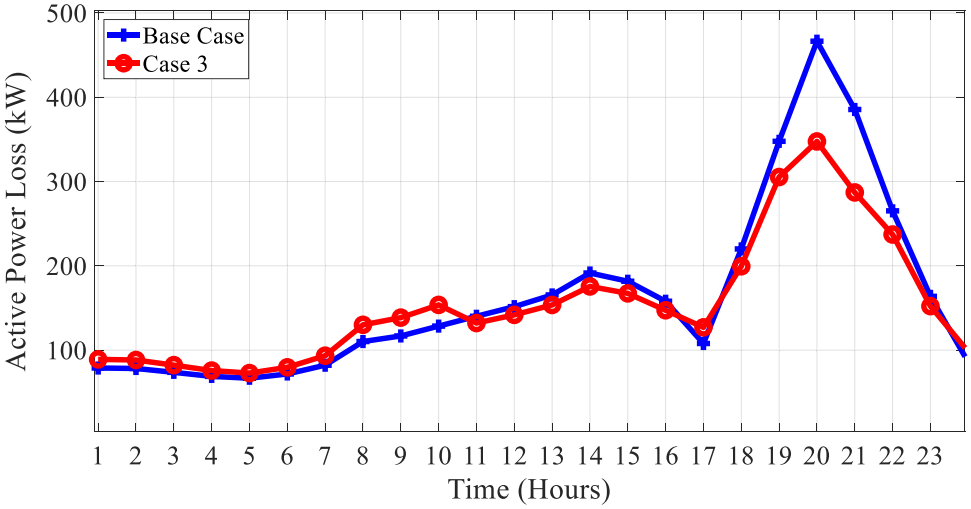


Fig. 9 Impact of 10% DR rate on active power losses

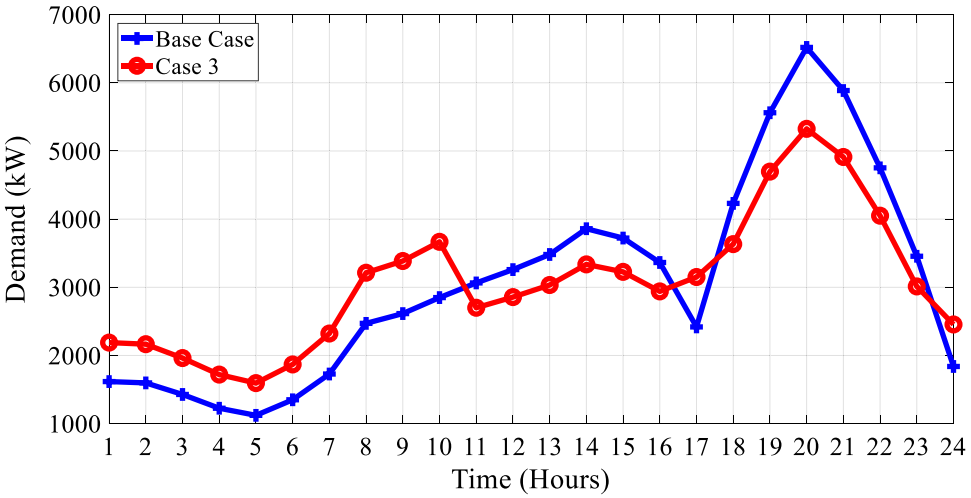


Fig. 10 Impact of 20% DR rate on demand pattern

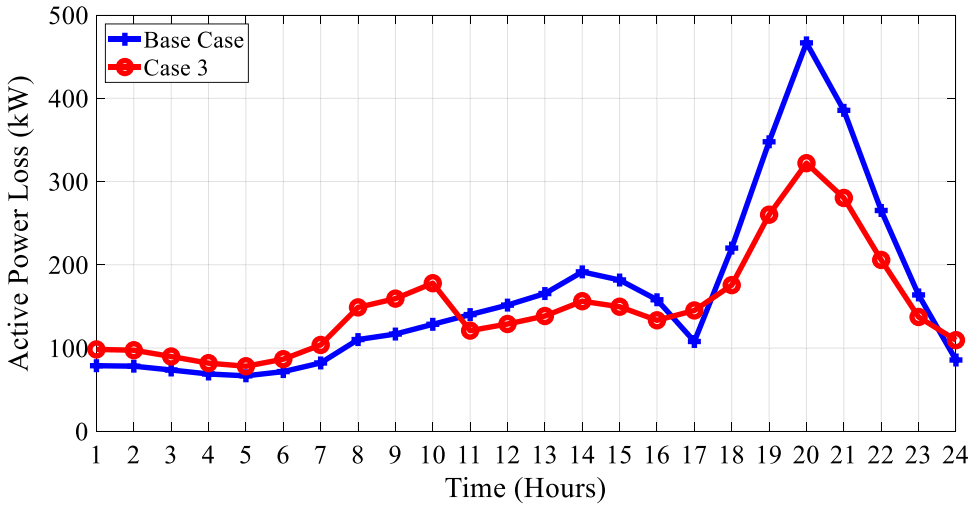


Fig. 11 Impact of 20% DR rate on active power losses

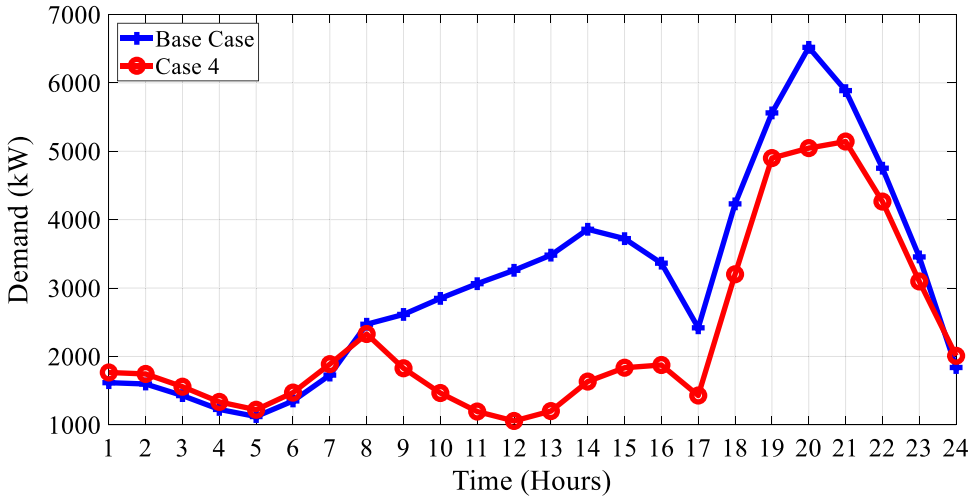


Fig. 12 Impact of DG and 10% DR rate on demand pattern

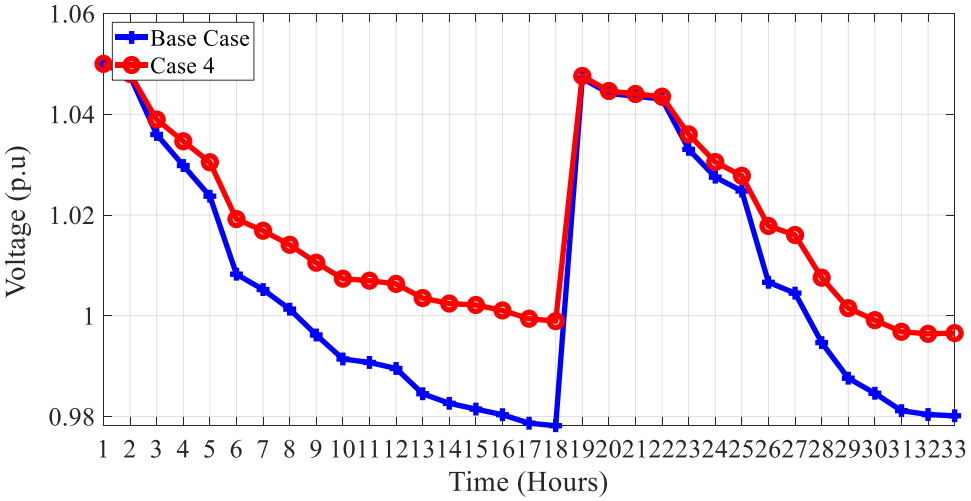


Fig. 13 Impact of DG and 10% DR rate on voltage pattern

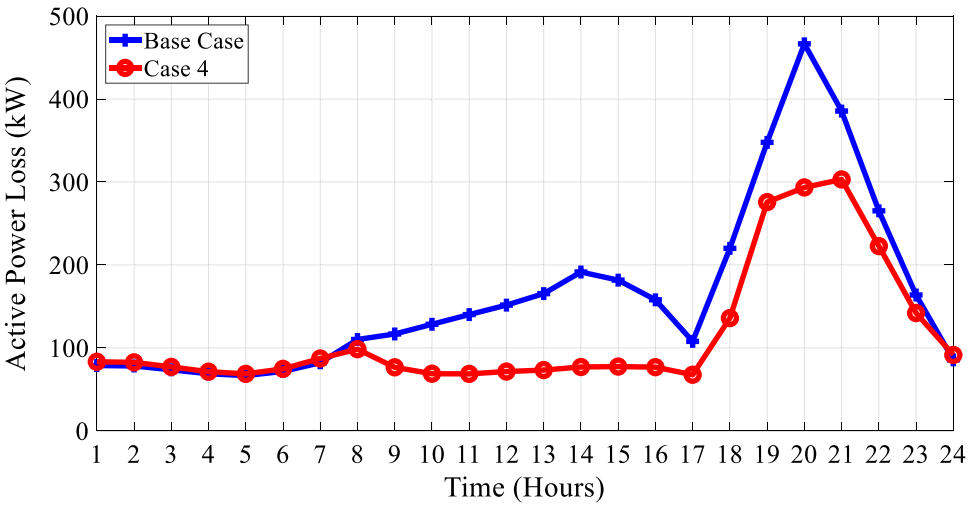


Fig. 14 Impact of DG and 10% DR rate on active power losses

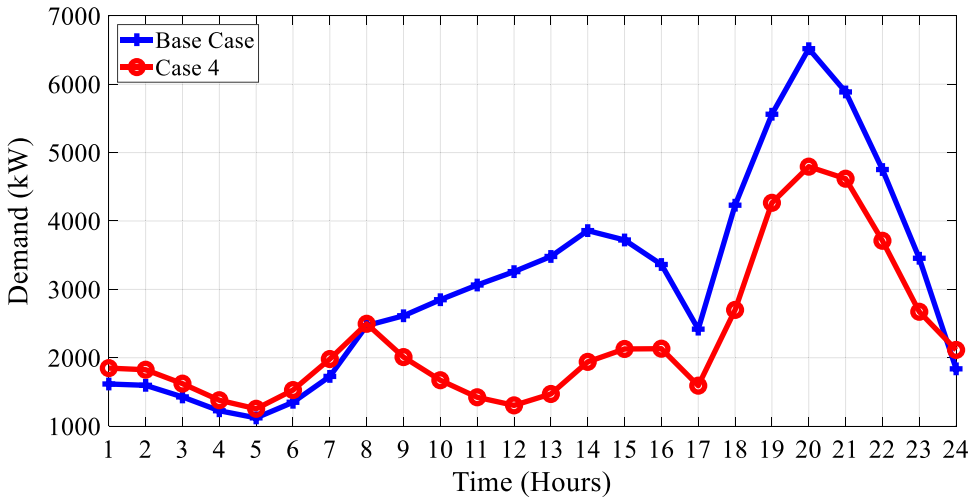


Fig. 15 Impact of DG and 20% DR rate on demand pattern

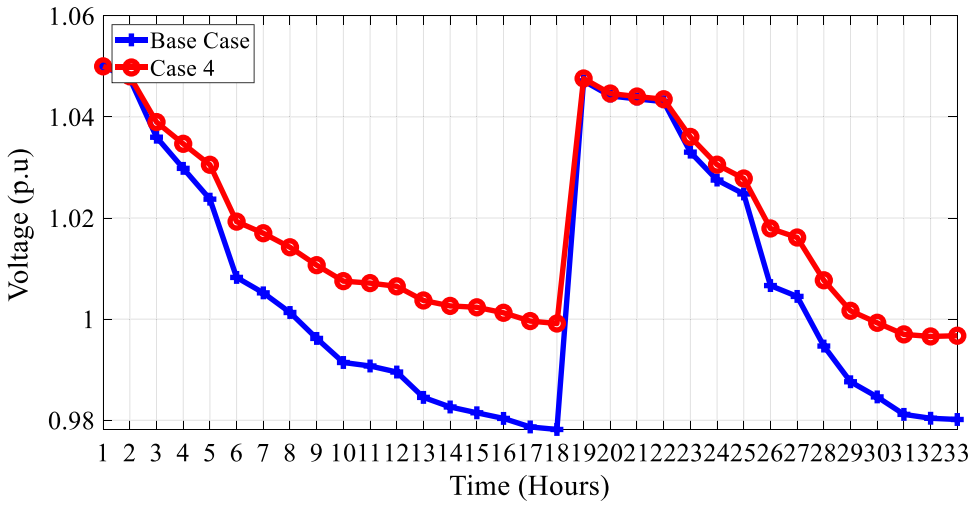


Fig. 16 Impact of DG and 20% DR rate on voltage pattern

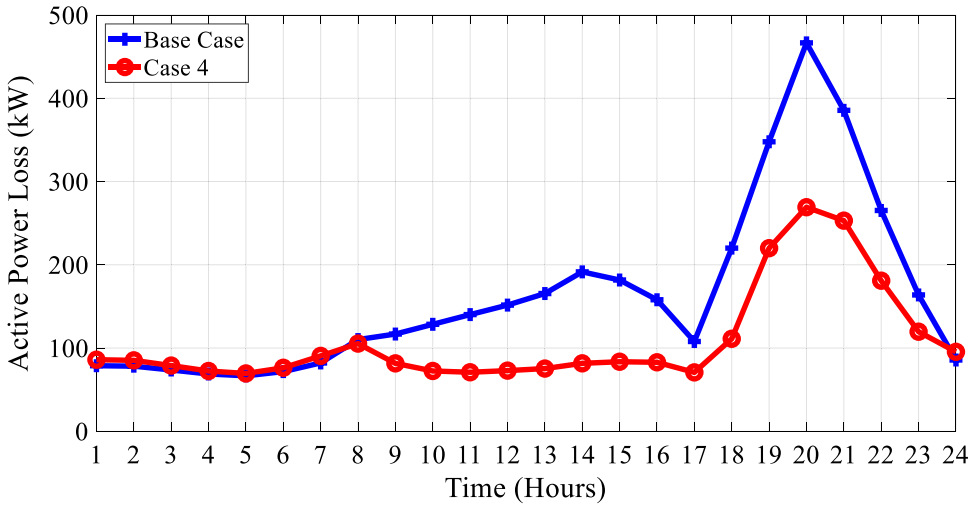


Fig. 17 Impact of DG and 20% DR rate on active power losses

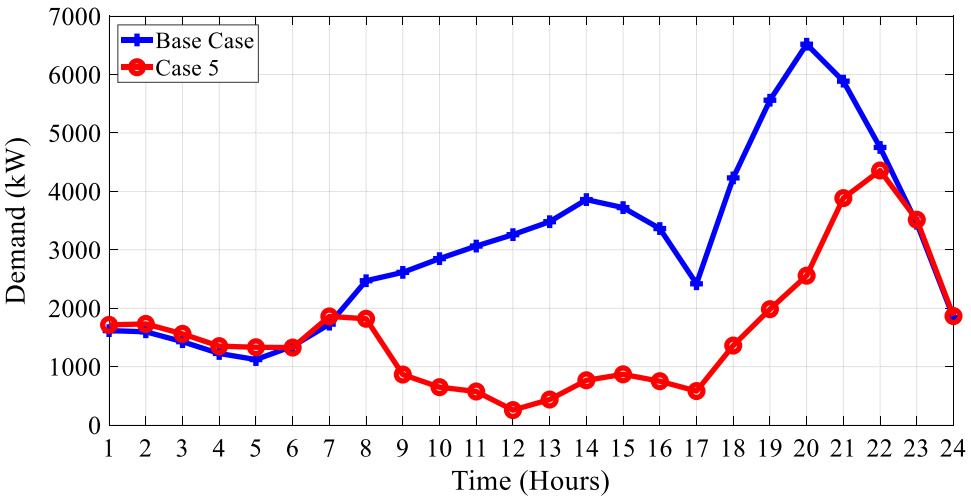


Fig. 18 Impact of DG and BESS on demand pattern

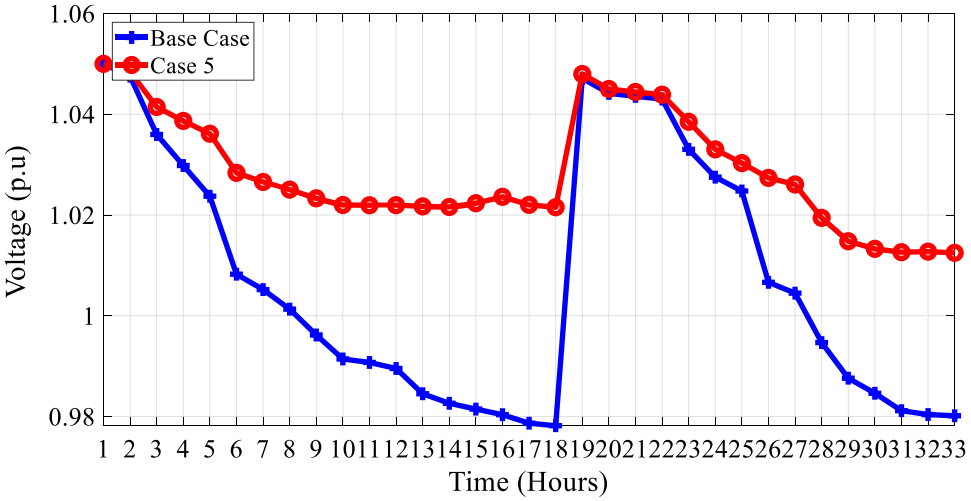


Fig. 19 Impact of DG and BESS on voltage pattern

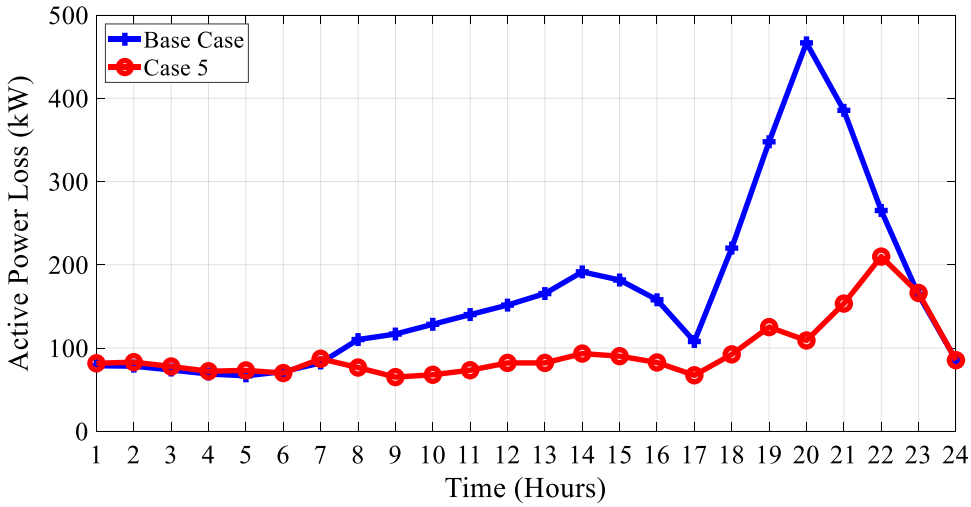


Fig. 20 Impact of DG and BESS on active power losses

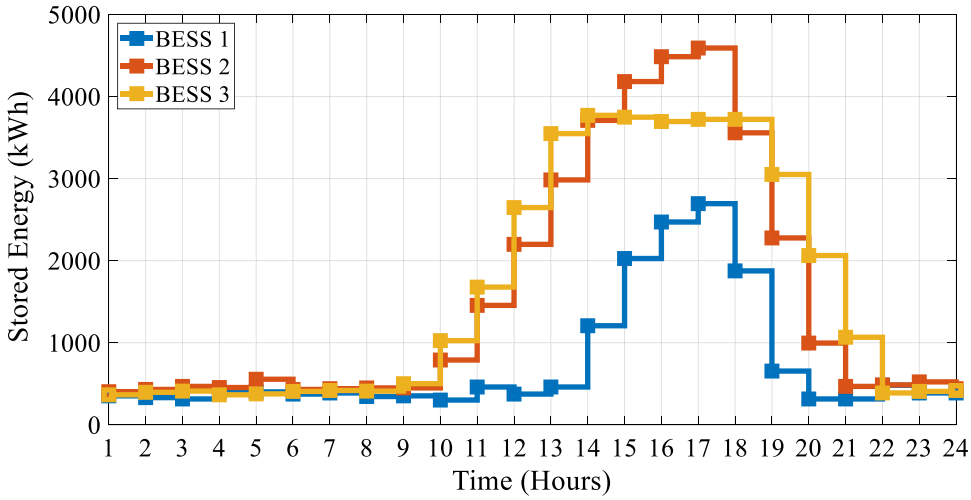


Fig. 21 BESS energy storage (DG + BESS)

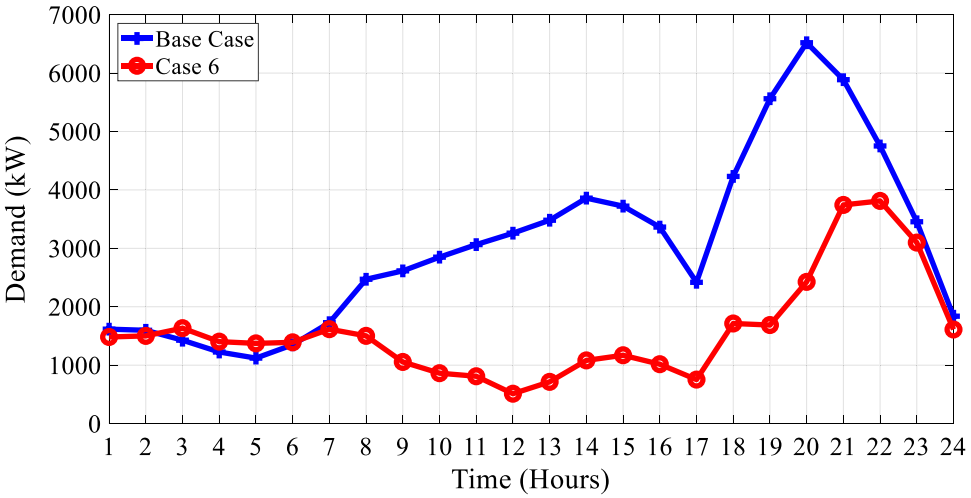


Fig. 22 Impact of DG, BESS and, 10% DR rate on demand pattern

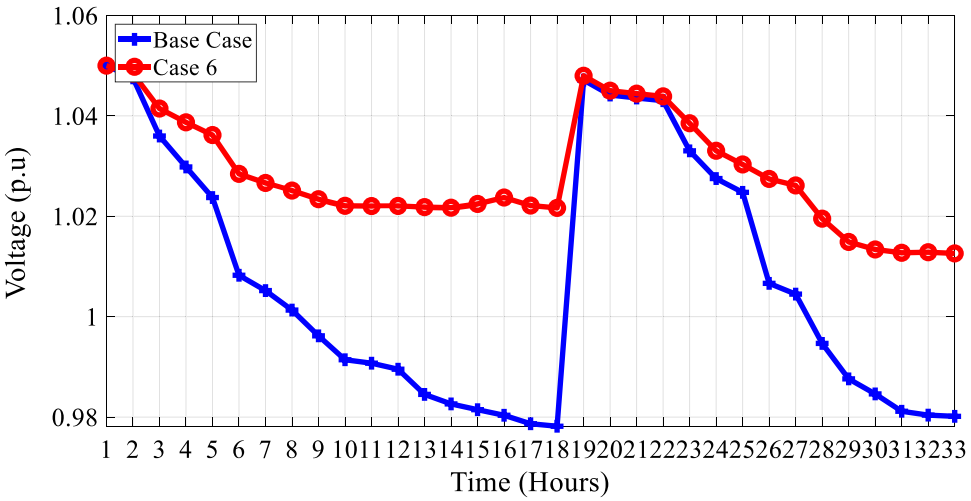


Fig. 23 Impact of DG, BESS and, 10% DR rate on voltage pattern

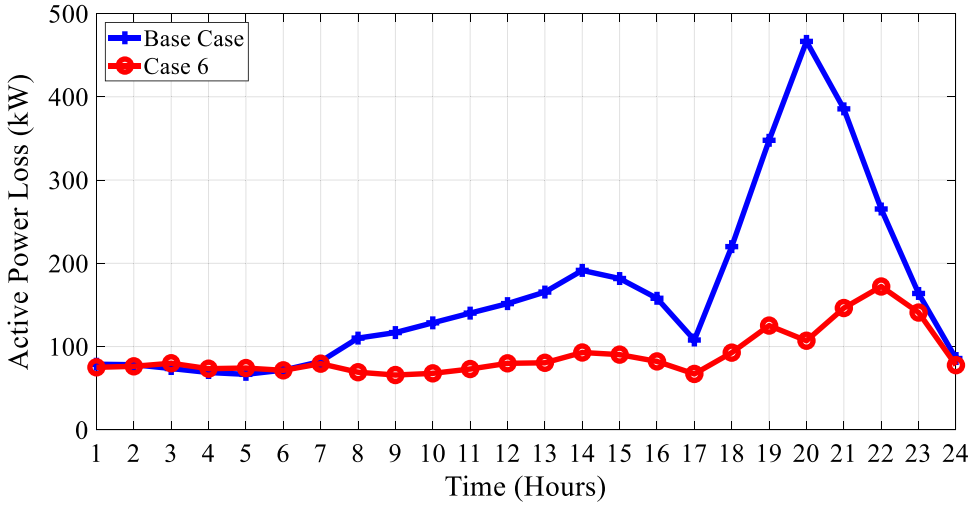


Fig. 24 Impact of DG, BESS and, 10% DR rate on active power losses

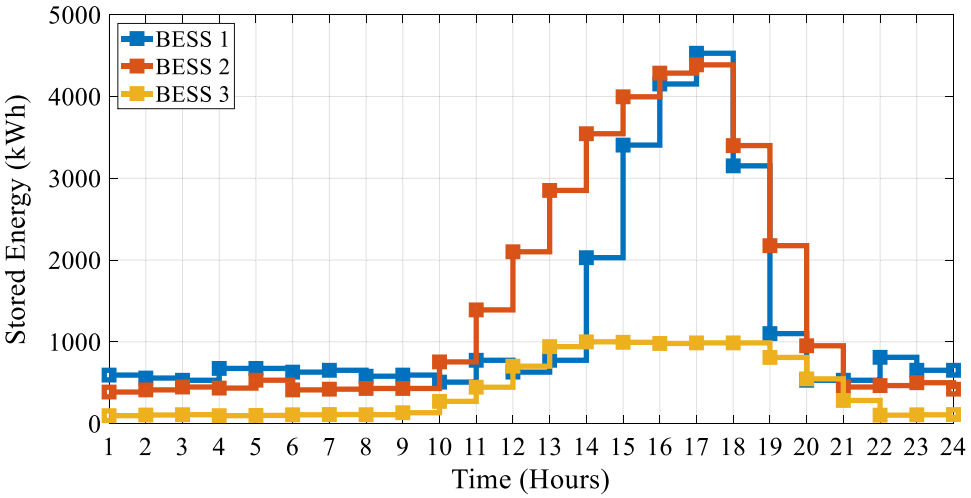


Fig. 25 BESS energy storage (DG, BESS and, 10% DR rate)

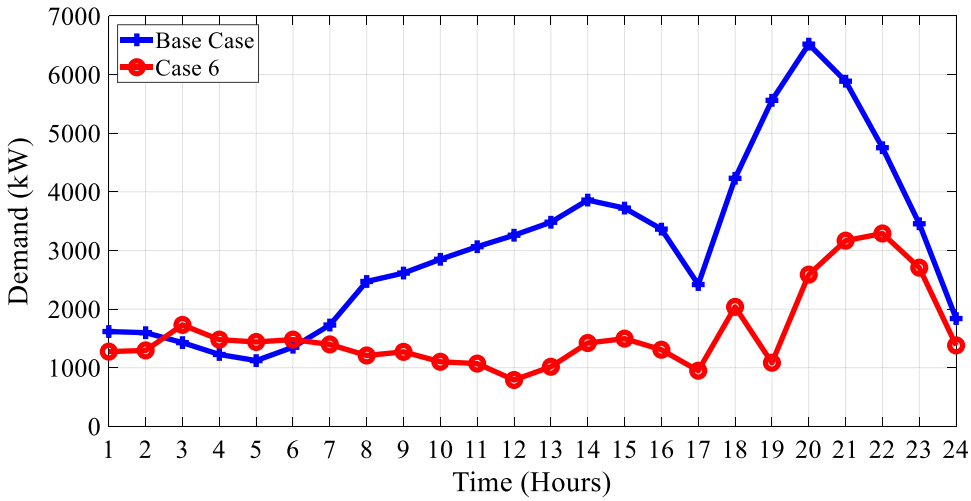


Fig. 26 Impact of DG, BESS, and 20% DR rate on demand pattern

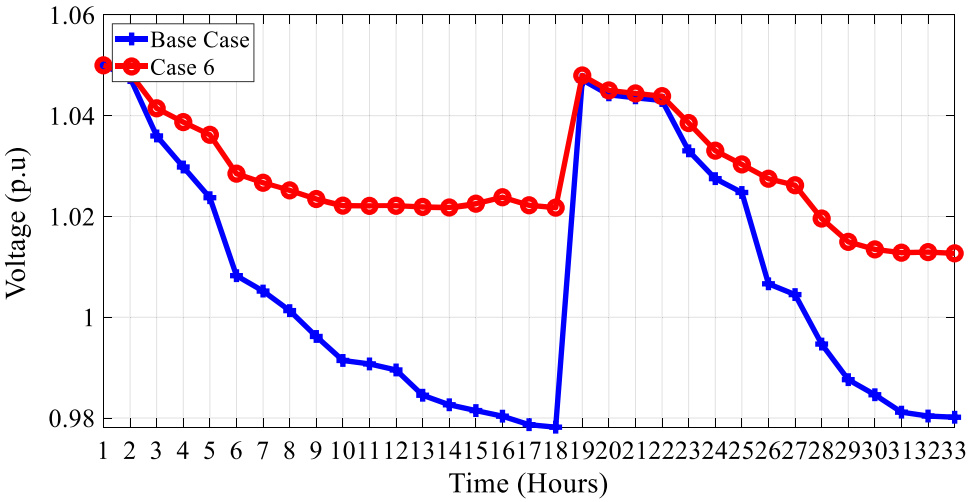


Fig. 27 Impact of DG, BESS, and20% DR rate on voltage pattern

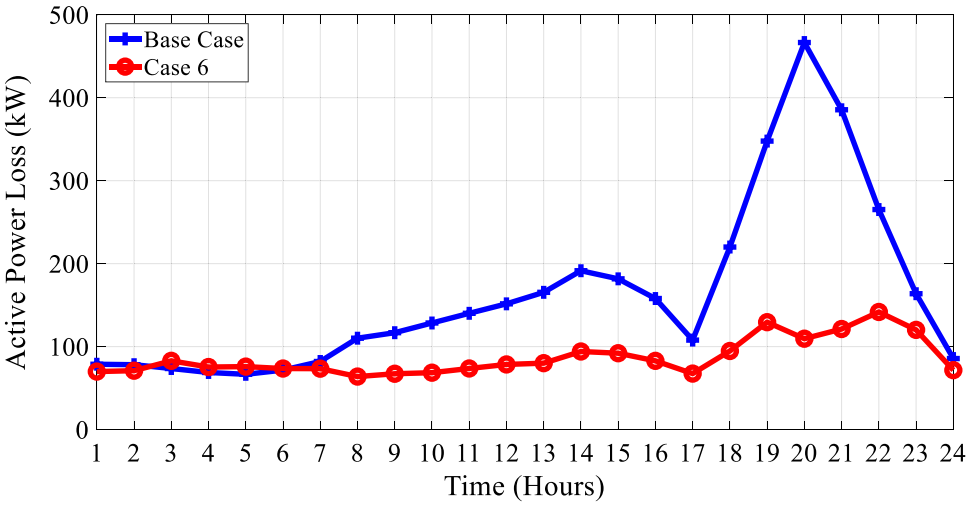


Fig. 28 Impact of DG, BESS and, 20% DR rate on active power losses

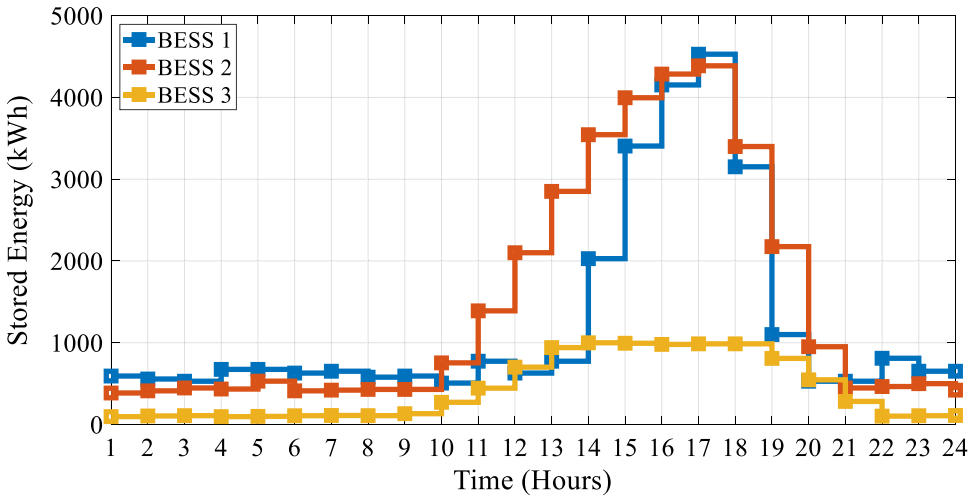


Fig. 29 BESS energy storage (DG, BESS and, 20% DR rate)

power losses, and BESS energy storage. The CO_2 emission is reduced by 32.71% in comparison with the base case.

5 Conclusions

In the context of future DN, three crucial elements are renewable energy utilization, demand DR planning, and the energy storage system. To maximize the technical, financial, and environmental benefits of DN, a well-coordinated optimization approach is proposed in this paper. The methodology focuses on efficiently planning and executing upgrades to the DN. This multilevel optimization system optimizes the coordination of SPV penetration, DR scheduling, and BESS allocation. At level-1, the optimal sizes and locations of DGs and BESS are determined, considering various constraints. At level-2, the DR program is implemented, coordinating with the outcomes from level-1 and dispatching BESS to maintain the balance between renewable energy generation and load demand. To demonstrate the effectiveness of the proposed technique, the IEEE 33 test system is employed. The harmonious integration of DGs, BESS, and DR is explored to achieve the highest possible DN performance. Notably, the performance of DN is significantly impacted by the coordination of different DR rates. In summary, this research highlights the importance of a coordinated approach in future DN, emphasizing the benefits of renewable energy, DR, and energy storage integration.

The implementation of the proposed approach demonstrates that the most substantial reduction in CO_2 emissions (32.71%) is achieved through the coordination of DG, BESS, and a 20% DR rate. Simultaneously, the maximum penetration of renewable energy (89.96%) is attained through the optimal allocation of DG in the coordination with BESS (Figs. 30 and 31).

i. In terms of performance, DGs prove to be highly effective in reducing annual energy losses. However, they do not consistently improve the load profile, and in some cases, may even have a negative impact on it. Furthermore, increased DG penetration can lead to voltage level surges and potential reverse power flow to the grid, setting a limit on their integration into the DS.
ii. Besides improving the load profile and voltage profile while minimizing energy losses, the inclusion of BESS enables higher DG penetration.
iii. Implementing DR successfully levels the load profile, reducing the gap between peak and minimum demand. This eases stress on the system and provides additional benefits, such as reduced reliance on BESS. Consequently, lower BESS deployment leads to a lower DG penetration. Conversely, if PV and BESS penetration is expected to be lower, a higher DR rate offers an efficient means to enhance DS efficiency.
iv. It can be deduced that a higher DR rate relative to optimal results leads to lower DG penetration. A decreased DG power dispatch results in lower energy loss minimization, and a high DR value may negatively impact the system. Hence, a coordinated approach among DGs, BESS, and DR within the DS is crucial.
v. While integrating a Solar PV system into the network can lead to a reduction in CO_2 emissions, the most favorable outcomes in terms of both CO_2 emissions and power quality parameters are achieved when implementing the proposed framework.
vi. The DR scheme exhibits a minimal impact on CO_2 emission reduction when implemented independently. However, its effectiveness significantly improves when integrated into a coordinated system with renewable DG and BESS.

Fig. 30 Total CO_2 emission/day

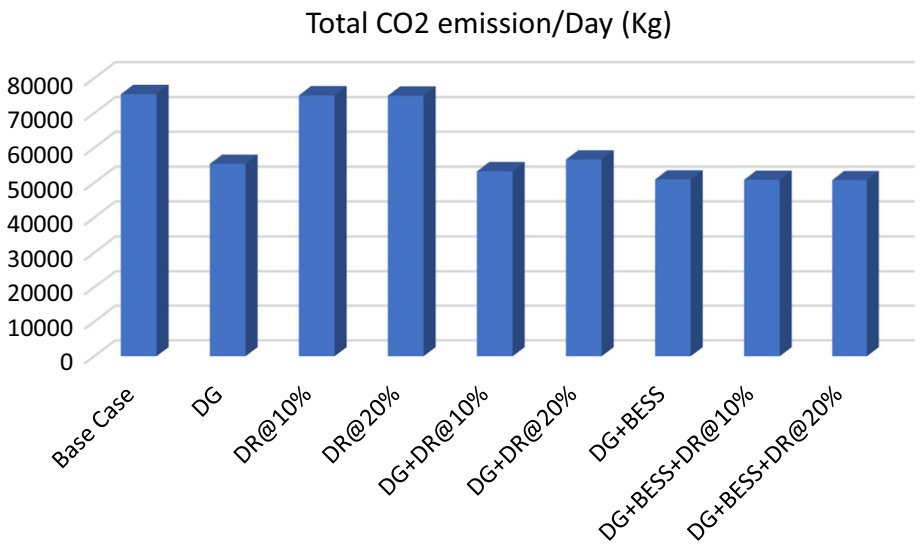
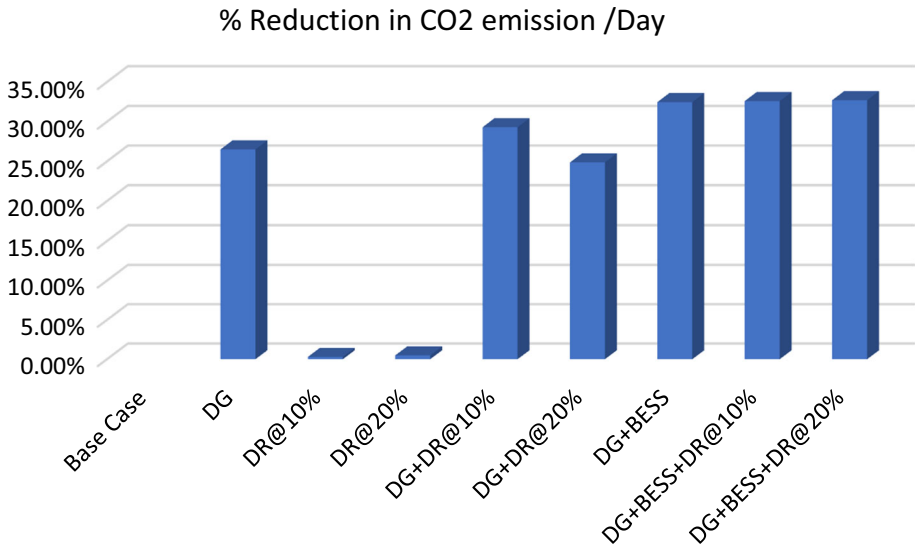


Fig. 31 Reduction in CO_2 emission by CPP



Author contributions Credit Author Statement Vivek Saxena: Conceptualization, Methodology, Software, Data curation, Writing- Original draft preparation, Visualization, Investigation, Validation, Writing- Reviewing and Editing. Narendra Kumar: Supervision Uma Nangia: Supervision

Funding This research received no external funding.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Saxena V, Kumar N, Nangia U (2007) Smart grid: a sustainable smart approach. In: Journal of physics: conference series, p 012042
2. Parsa Moghaddam M, Abdollahi A, Rashidinejad M (2011) Flexible demand response programs modeling in competitive electricity markets. Appl Energy 88(9):3257–3269
3. Medina J, Muller N, Roytelman I (2010) Demand response and distribution grid operations: opportunities and challenges. IEEE Trans Smart Grid 1(2):193–198
4. Saxena V, Kumar N, Nangia U (2021) Analysis of smart electricity grid framework unified with renewably distributed generation. In: Advances in smart communication and imaging systems. Lecture Notes in Electrical Engineering, pp 735–751
5. Hung DQ, Mithulananthan N (2013) Multiple distributed generator placement in primary distribution networks for loss reduction. IEEE Trans Ind Electr 60(4):1700–1708
6. Esmaili M (2013) Placement of minimum distributed generation units observing power losses and voltage stability with network constraints. IET Gener Transm Distrib 7(8):813–821
7. Georgilakis PS, Hatziargyriou ND (2013) Optimal distributed generation placement in power distribution networks: models, methods, and future research. IEEE Trans Power Syst 28(3):3420–3428

8. Saxena V, Kumar N, Nangia U (2022) Recent trends in the optimization of renewable distributed generation: a review. Ing Investig 42(3):e97702
9. Hashemi S, Stergaard J, Yang G (2014) A scenario-based approach for energy storage capacity determination in LV grids with high PV penetration. IEEE Trans Smart Grid 5(3):1514–1522
10. Zheng Y, Dong ZY, Luo FJ, Meng K, Qiu J, Wong KP (2014) Optimal allocation of energy storage system for risk mitigation of discos with high renewable penetrations. IEEE Trans Power Syst 29(1):212–220
11. Zheng Y, Hill DJ, Dong ZY (2017) Multi-agent optimal allocation of energy storage systems in distribution systems. IEEE Trans Sustain Energy 8(4):1715–1725
12. Barcellona S, Piegari L, Musolino V, Ballif C (2018) Economic viability for residential battery storage systems in grid-connected PV plants. IET Renew Power Gener 12(2):135–142
13. Paliwal NK, Singh AK, Singh NK, Kumar P (2018) Optimal sizing and operation of battery storage for economic operation of hybrid power system using artificial bee colony algorithm. Int Trans Electr Energy Syst 29(1):e2685
14. Nick M, Cherkaoui R, Paolone M (2018) Optimal planning of distributed energy storage systems in active distribution networks embedding grid reconfiguration. IEEE Trans Power Syst 33(2):1577–1590
15. Banguero E, Correcher A, Navarro P, Morant F, Aristizabal A (2018) A review on battery charging and discharging control strategies. Appl Renew Energy Syst Energies 11(4):1021
16. Safdarian A, Fotuhi-Firuzabad M, Lehtonen M (2016) Benefits of demand response on operation of distribution networks: a case study. IEEE Syst J 10(1):189–197
17. Cresta M, Gatta FM, Geri A, Maccioni M, Mantineo A, Paulucci M (2015) Optimal operation of a low-voltage distribution network with renewable distributed generation by NaS battery and demand response strategy: a case study in a trial site. IET Renew Power Gener 9(6):549–556
18. Roscoe AJ, Ault G (2010) Supporting high penetrations of renewable generation via implementation of real-time electricity pricing and demand response. IET Renew Power Gener 4(4):369–382
19. Sharma S, Niazi KR, Verma K, Rawat T (2020) Coordination of different DGs, BESS and demand response for multi-objective optimization of distribution network with special reference to Indian power sector. Int J Electr Power Energy Syst 121:106074
20. Sharma S, Niazi KR, Verma K, Rawat T (2020) A bi-level optimization framework for investment planning of distributed generation resources in coordination with demand response. Energy Sources Part A Recovery Util Environ Effects. https://doi.org/10.1080/15567036.2020.1758248
21. Sharma S, Niazi KR, Verma K, Rawat T (2020) Impact of battery energy storage, controllable load and network reconfiguration on contemporary distribution network under uncertain environment. IET Gener Transm Distrib 14(21):4719–4727
22. Yang C, Wu Z, Li X, Fars A (2024) Risk-constrained stochastic scheduling for energy hub: integrating renewables, demand response, and electric vehicles. Energy 288:129680
23. Shen Y, Xie J, He T, Yao L, Xiao Y (2023) CEEMD-fuzzy control energy management of hybrid energy storage systems in electric vehicles. IEEE Trans Energy Convers. https://doi.org/10.1109/TEC.2023.3306804
24. Li R, Xu D, Tian H, Zhu Y (2023) Multi-objective study and optimization of a solar-boosted geothermal flash cycle integrated into an innovative combined power and desalinated water production process: application of a case study. Energy 282:128706
25. Cao B, Dong W, Lv Z, Gu Y, Singh S, Kumar P (2020) Hybrid microgrid many-objective sizing optimization with fuzzy decision. IEEE Trans Fuzzy Syst 28(11):2702–2710
26. Tian H, Li R, Salah B, Thinh P-H (2023) Bi-objective optimization and environmental assessment of SOFC-based cogeneration system: performance evaluation with various organic fluids. Process Saf Environ Prot 178:311–330
27. Li P, Hu J, Qiu L, Zhao Y, Ghosh BK (2022) A distributed economic dispatch strategy for power-water networks. IEEE Trans Control Netw Syst 9(1):356–366
28. Xiao T, Lin Z, Liu C, Liu L, Li Q (2023) Integration of desalination and energy conversion in a thermo-osmotic system using low-grade heat: performance analysis and techno-economic evaluation. Appl Therm Eng 223:120039
29. Duan Y, Zhao Y, Hu J (2023) An initialization-free distributed algorithm for dynamic economic dispatch problems in microgrid: modeling, optimization and analysis. Sustain Energy Grids Netw 34:101004
30. Chen H, Wu H, Kan T, Zhang J, Li H (2023) Low-carbon economic dispatch of integrated energy system containing electric hydrogen production based on VMD-GRU short-term wind power prediction. Int J Electr Power Energy Syst 154:109420
31. Shirkhani M, Tavoosi J, Danyali S, Khosravi Sarvenoee A (2023) A review on microgrid decentralized energy/voltage control structures and methods. Energy Rep 10:368–380
32. Li H-M, Li G-X, Li L, Wu J-Z, Yao Z-P, Zhang T (2023) Combustion characteristics and concentration measurement of ADN-based liquid propellant with electrical ignition method in a combustion chamber. Fuel 344:128142
33. Liu K, Sheng W, Li Z, Liu F, Liu Q, Huang Y, Li Y (2023) An energy optimal schedule method for distribution network considering the access of distributed generation and energy storage. IET Gener Transm Distrib 17:2996–3015
34. Nosratabadi SM, Hooshmand RA, Gholipour E, Parastegari M (2016) A new simultaneous placement of distributed generation and demand response resources to determine virtual power plant. Int Trans Electr Energy Syst 26(5):1103–1120
35. Eid C, Koliou E, Valles M, Reneses J, Hakvoort R (2016) Time-based pricing and electricity demand response: existing barriers and next steps. Util Policy 40:15–25
36. Zhang C, Xu Y, Dong ZY, Wong KP (2017) Robust coordination of distributed generation and price-based demand response in microgrids. IEEE Trans Smart Grid 9(5):4236–4247
37. CEA Report (2023) https://cea.nic.in/wpcontent/uploads/baseline/2023/01/Approved_report_emission__2021_22.pdf
38. Rajput SK, Dheer DK (2022) Integration of 100-kWp PV with low-voltage distribution power system in composite climate: performance and energy metrics analysis. Int J Ambient Energy 43:8176–8192
39. Meena NK, Parashar S, Swarnkar A, Gupta N, Niazi KR (2018) Improved elephant herding optimization for multiobjective DER accommodation in distribution systems. IEEE Trans Ind Inf 14(3):1029–1039

40. Kanwar N, Gupta N, Niazi KR, Swarnkar A (2015) Improved meta-heuristic techniques for simultaneous capacitor and DG allocation in radial distribution networks. Int J Electr Power Energy Syst 73:653–664
41. Saxena V, Kumar N, Nangia U (2022) An impact assessment of distributed generation in distribution network. In: Artificial intelligence and sustainable computing, pp 33–46
42. Safdarian A, Degefa MZ, Lehtonen M, Fotuhi-Firuzabad M (2014) Distribution network reliability improvements in presence of demand response. IET Gener Transm Distrib 8(12):2027–2203
43. Saxena V, Kumar N, Nangia U (2022) An extensive data-based assessment of optimization techniques for distributed generation allocation: conventional to modern. Arch Comput Methods Eng 1–27
44. Baran ME, Wu FF (1989) Network reconfiguration in distribution systems for loss reduction and load balancing. IEEE Trans Power Deliv 4(2):1401–1407
45. Mazidi M, Zakariazadeh A, Jadid S, Siano P (2014) Integrated scheduling of renewable generation and demand response programs in a microgrid. Energy Conver Manag 86:1118–1127

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.

Communication independent Islanding Detection technique utilizing Wavelet Packet Transform

Samavedam Swetha, Vishal Verma
Electrical Engineering Department
Delhi Technological University
Shahbad Daulatpur Village, Rohini, New Delhi, Delhi 110042
swethasamavedam@gmail.com, vishalverma@dce.ac.in

2023 IEEE International Conference on Power Electronics, Smart Grid, and Renewable Energy (PESGRE) | 979-8-3503-1057-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/PESGRE58662.2023.10404137

***Abstract*— Islanding detection becomes a necessity when the DERs units are required to continue their generation even after the islanding of the µG from the grid. This paper investigates the effectiveness of the active islanding detection technique using 9th harmonic pilots' injection and its detection with wavelet packet decomposition at each node in autonomous mode without any communication equipment by extracting the features that can indicate the occurrence of islanding and outage of DERs connected in the µG. The proposed method is evaluated through simulation studies on a typical microgrid system having many DERs connected, and the results demonstrate the effectiveness of the method in detecting islanding events in respect of both grid and DERs with high accuracy and reliability making the power system more secure and resilient.**

Keywords—Islanding detection, Wavelet Packet Transform, harmonic current injection, distributed energy sources.

I. INTRODUCTION

To build a more sustainable and resilient energy system that benefits both the economy and the environment, the shift towards Renewable Energy Sources(RESs) has become increasingly crucial. Firstly, RESs such as solar and wind power[1] generate electricity without releasing harmful greenhouse gases, unlike fossil fuels, thereby climate change is mitigated through reduction in carbon footprints in terms of energy generation. Additionally, the affordability of renewable energy sources is rising in comparison to conventional fossil fuels, rendering them economically feasible. RES employment offers a decentralised energy system in the form of a µG consisting of Distributed Energy Resources (DERs) that can increase energy reliability, stability of the grid, reduction in the transmission losses, improvement in environmental sustainability, and reduction of stress on centralized generation [2].

However, integrating DERs with the µG pose several challenges owing to the different fault characteristics and control strategies of the DERs than that of the µG [3]. Detection of islanding conditions of the grid and the DERs quickly and accurately is of significant concern since, in such situations, the DERs left connected in the absence of the grid continue to provide power to the connected load with at least one DER acting as a voltage source in the µG, even when the utility's power supply is no longer available for maintenance of the energy balance [4]. This imposes multiple operational constraints such as voltage instability, transients, and power quality disturbances, etc., which can affect the reliability and efficiency of the µG adhering to the existing standards such as IEEE 1547 – 2018 [5], UL 1741 [6], and IEC – 62116. The efficiency of islanding detection techniques is gauged in terms of their fastness in error detection, power quality, Non – Detection Zone (NDZ), and their undeterred operation amidst cases, where there are multiple DERs connected in the µG. This calls for study on various islanding detection techniques. Figure 1 illustrates the categorization of the islanding detection techniques into remote, local, and intelligent classifier-based methods.

Passive techniques rely on measuring electrical quantities that reflect the inherent characteristics of islanding[7] – [12]. In contrast the active techniques involve intentional introduction of distortion in the system and studying its effects for detection purposes[13] – [17]. Hybrid techniques represent a fusion of both passive and active approaches[18] – [23], which involves signal processing techniques combined with Active islanding detection techniques, thus providing less polluting, accurate and fast detection of grid islanding detection. In order to decrease the non-detection zone (NDZ) of passive methods, contemporary signal processing techniques such as the Short-Time Fourier Transform (STFT) and wavelet transform (WT) are utilized. Although STFT utilises a time frequency window to localised transients, lacks in the ability to provide multiple resolutions due to its fixed window size[24]. As the frequency increases, more cycles are recommended within the fixed window causing the individual frequency components to become less localized. This limitation is overcome by the wavelet transform, which incorporates the principles of STFT and multi resolution techniques. Unlike STFT, the WT employs variable windows for signal analysis. However, it solely captures the low-frequency range, neglecting valuable properties present in the high-frequency band that are relevant for revealing islanding conditions. Thus, leaving behind the scope of hybrid wavelet techniques which can be spread over the entire spectra to alleviate the shortfalls of WT based detection schemes.

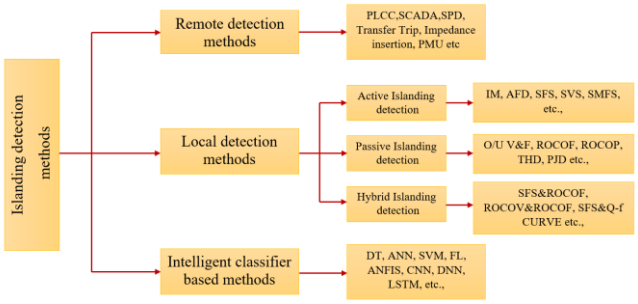


Fig 1: Classification of Islanding detection methods

The utility grid operates as voltage source, operating in grid – forming mode (GFM), giving opportunity for the other inverters to operate in the grid supporting mode (GSM) [4]. In the event of the utility system disconnection during the master-slave operation of the µG, a specific distributed energy resource (DER) might be required to operate in grid-following mode (GFM), while other DERs (slaves) continue operating with maximum power point tracking (MPPT). Additionally,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:16:10 UTC from IEEE Xplore. Restrictions apply.

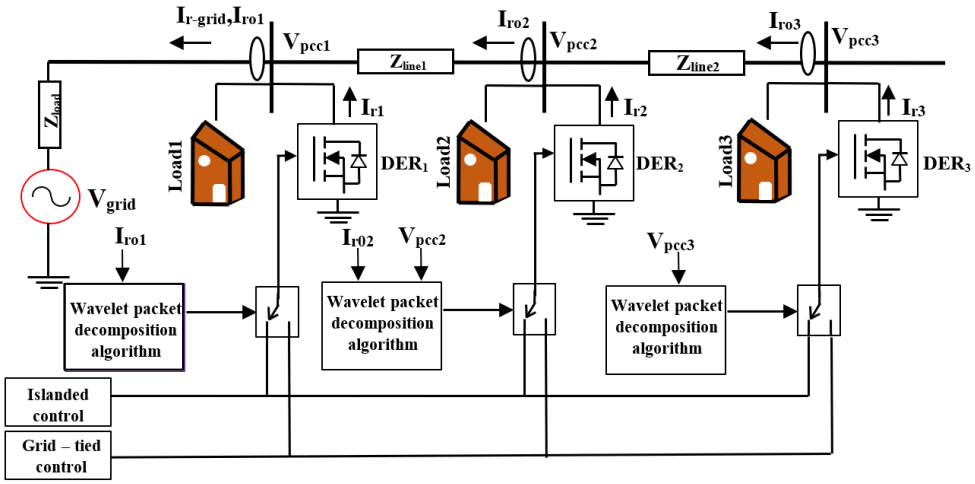


Fig 2: Microgrid system employed to check the effectiveness of the Wavelet packet decomposition algorithm for islanding detection.

certain DERs (slaves) may need to operate in a de-loaded mode, deviating from MPPT, in order to align with the power demand. A contingency situation may arise with the outage of one or more DER's in the islanded mode, which also need to be stabilized by the connected DERs, shifting their operation from de-loaded mode quickly towards MPPT mode by reducing the de-loading factor. Since, most inverters operate in the Grid-Supportive Mode (GSM), injecting current using Maximum PowerPoint Tracking (MPPT), they are often blind on the status of other DERs connected to the µG. This also calls for operational shift towards active grid with communication independent, immaculate control and coordination among the various DERs connected in the µG at every level.

This paper proposes an islanding detection method utilising DER tagging strategy that is reliable, effective, and fast when multiple DERs are present in a µG and detection is responsible not only for grid islanding but also for outage detection of DERs. The proposed method involves the injection of 9th harmonic pilots, which can quickly identify grid islanding, along with the outage of specific DERs. As an alternative, this paper proposes the use of Discrete Wavelet Packet Transform (WPT), dividing the entire time-frequency frame, while the WT is limited to analysing the low frequency band alone. This method overcomes sensitivity to noise, as in case of Fuzzy Inferencing based method of islanding detection. The WPT method is used at each node operates in autonomous mode to analyse and detect the specific patterns and frequency content in the 9th harmonic component of the voltage at Point of Common Coupling (PCC) in conjunction with the current in the line around each node, including the node which is transacting power with the grid interface. The injected balanced 9th harmonic currents are set to a fixed value and sum up vectorially at the grid interfaced node to a zero value under normal conditions of operation. Being equi-displaced in phase with respect to one another, it is likely to introduce minimal harmonic impact on account of their escape to mainly the grid by each DER.

II. ISLANDING DETECTION SCHEME

The system considered for analysis and design in this study is shown in Fig2. The configuration consists of three photovoltaic-fed inverter DERs, each with a maximum power rating of 4.2kW, connected to the 415V distribution grid, and injecting power at unity power factor through their respective PCCs in GSM. At each PCC node, there is a local load of 5.67kW. The utility grid is rated at an AC voltage of 415V, 50Hz, and lumped line impedance Z_{line1} and Z_{line2} to evaluate the proposed islanding detection algorithm efficiency. To approximate the behavior of the system, the line impedance between the two sources is considered resistive, and the voltage at PCC(V_{pcc}) of each DER is in approximately the same phase as the others due to minimum line reactance.

A. Implementation of Harmonics injection for islanding detection:

In the analysed configuration, which is a 3-phase 4-wire system, the current flowing through the neutral wire consists of the 3rd harmonic and its odd multiples. However, these harmonics balance out in the line, making it challenging to detect islanding conditions. To address this, a technique known as harmonic pilots can be utilized. By injecting harmonic pilots of the order 3n (where n = 1, 3, 5, 7, etc.) into the line, it becomes possible to enable effective islanding detection[26]. In this paper, 9th harmonic current and/or voltage harmonics are used as a unique signature for detecting islanding events since these are not commonly present in the power system under normal operating conditions but also because their frequency being high enough to be easily distinguished from the fundamental frequency, but low enough to avoid significant signal attenuation in the power system. The phase angle between the 9th harmonic current injected by the three DERs is calculated to be 120˚(360˚/n), where n is equal to 3, as the system is 3-phase. For this discussion, the focus will be on the single-phase current and voltage phasers, with the understanding that the same logic applies to the other two phases. The first DER, i.e., DER_1 injects current at the 9th harmonic at 0˚, while DER_2 injects current at 240˚ and DER_3 at 120˚ respectively. The sum of the three harmonic current phasors at the 9th harmonic adds up to zero, resulting in a net harmonic current of zero injected into the grid, as shown in Fig 3.

B. Suggested DG failure detection methodology:

The current phasors at the 9th harmonic injected by DER_1, DER_2 and DER_3 into one phase, say r-phase of the grid can be represented as I_{r1}, I_{r2} and I_{r3}. Similarly, I_{ro1}, I_{ro2}, and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:16:10 UTC from IEEE Xplore. Restrictions apply.

I_{ro3}, are the line currents of one phase at the 9th harmonic in the sections of the μG, while I_{r_grid} denotes the current flowing into the grid.

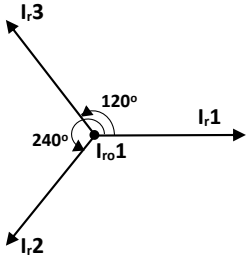


Fig 3: Proposed 9th harmonic pilots' injection technique

The relation between the above phasors can be approximated as follows:

$$I_{ro3} = I_{r3}$$
$$I_{ro2} = I_{ro3} + I_{r2}$$
$$I_{ro1} = I_{ro2} + I_{r1}$$
$$I_{ro1} = I_{r-grid}$$

Similarly, the voltages at the V_{pcc} of DER_1, DER_2 and DER_3 are depicted as V_{pcc1}, V_{pcc2}, and V_{pcc3} respectively. Under normal condition of operation, V_{pcc1} is zero, because the net harmonic 9th harmonic current, I_{ro1} flowing into the grid is zero. The relation between the PCC voltages and currents injected by each DER are as follows:

$$V_{pcc2} = I_{ro2} * (Z_{line1} + Z_{line2})$$
$$V_{pcc3} = (I_{ro2} * (Z_{line1} + Z_{line2})) + (I_{ro3} * Z_{line2})$$

In the event of a DER being isolated from the μG, the DERs upstream of the disconnected DER would detect the islanding by analysing the 9th harmonic current. Since the grid offers zero impedance to the 9th harmonic currents, the current only adds up in the forward direction. The DERs downstream of the isolated DER would detect the islanding by monitoring the 9th harmonic V_{pcc}, as the voltage phasors would not add up to zero due to absence of the current phasor that was previously being injected by the isolated DER. Consider a case where DER2 disconnects from the μG. The DERs upstream of the disconnected DER, i.e., DER_1 would know about it by observing the change in current phasor, while the DER's downstream, i.e., DER_3 would know of its disconnection by analysing the voltage phasor at its PCC. This method of islanding detection creates an opportunity for implementation of communication system failure proof islanding detection infrastructure, which is more reliable, cost – effective and free from communication delays, thereby reducing the time taken to implement the corrective measures.

C. Application of wavelet packet transform for islanding detection:

Wavelet packet decomposition is a useful signal processing tool that helps in islanding detection owing to its ability to effectively detect changes induced in the high frequency components of voltage and current signals post occurrence of islanding event. Compared to Fourier-based transforms such as STFT, FFT, and DFT, the wavelet transform offers several advantages. The main reason for this is that Fourier-based transforms use a fixed window size, whereas wavelet transforms allow for variable window sizes. As a result, wavelet transforms maintain both time and frequency resolutions without compromising either. Moreover, the computational capability of wavelet packet transform (WPT) to effectively capture time and frequency details of low and high frequencies using long and short windows, respectively, can be advantageous for precise islanding detection. This is particularly relevant in cases where the grid exhibits multiple current and voltage harmonic pilots.

The current and voltages at PCC of each DERs are analyzed by decomposing into its wavelet packets at various scales and frequencies using the Tree Decomposition algorithm. Sampling frequency($f_{sampling}$) of 10KHz has been used to sample the currents and voltages of the proposed system to demonstrate the efficacy of the proposed algorithm. According to Nyquist sampling theorem, the topmost frequency signal of the binary tree is $f_{sampling}/2$, i.e., 5KHz. This signal is split into two sub bands: the approximation coefficients(low pass filtered signal), i.e., signal in the range 0-2.5 KHz and the detail coefficients(high pass filtered signal) i.e., signal in the range 2.5-5 KHz. This tree decomposition method is further recursively applied to the splitting process to update the approximation coefficients until the desired number of decomposition levels is reached, where the tree is pruned to remove the unnecessary sub bands, such that 50 Hz in 450 Hz signals are distinctively separated as shown in Fig 4. The simplicity in the tree structure as shown in Fig 4 allows for accurate and detailed analysis compared to the Fourier-based methods.

Thus, 5-level binary tree-decomposition has been implemented which distinctively identifies the current and voltage signatures of 50Hz and 450 Hz pilots at (5,0) or 31 and (5,2) or 33 sub-bands respectively. By observing the 450Hz current and/or voltage pilots' signature at the respective PCC, information regarding islanding of other DERs in a network can be known, which therefore will help it to manage its power generation capability to meet the load demand. Islanding of different DERs in the system produce unique 450Hz current and/or voltage signatures, which are analyzed based on the location of the DER with respect to the disconnected DER.

Consider the case when DER_2 is disconnected from the grid. Disconnection of DER_2 is known to DER_1 by analyzing the 450Hz current signature of I_{ro1} at its PCC, since it being a upstream DER. Similarly, DER_3 would know of disconnection of DER_2 by analyzing the 450Hz voltage signature of V_{pcc3} at its PCC, since it being a downstream DER in the system. If both DER_2 and DER_3 are disconnected, then the unique current signature of I_{ro1} at the PCC of DER_1 can be used to infer the islanding.

III. SIMULATION RESULTS

To support the theoretical analysis of the proposed islanding detection technique, a simulation study was implemented in MATLAB/SIMULINK environment on the proposed microgrid system. The microgrid consists of three DERs with the capacity of 4.2 kW supplying a 5.67 kW load connected to each PCC. The line resistances, represented by Z_{line1} and Z_{line2} are 0.467Ω and 0.783Ω, respectively which corresponds to feeder line distances off 1.7 km and 2.5 km. The line is assumed to be resistive with a resistance to reactance ratio greater than 7. The grid is modelled at 415V and 50Hz. The connection and disconnection of each DG to the grid is implemented through circuit breakers. The wavelet

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:16:10 UTC from IEEE Xplore. Restrictions apply.

packet decomposition is implemented on signals sampled at 10KHz.

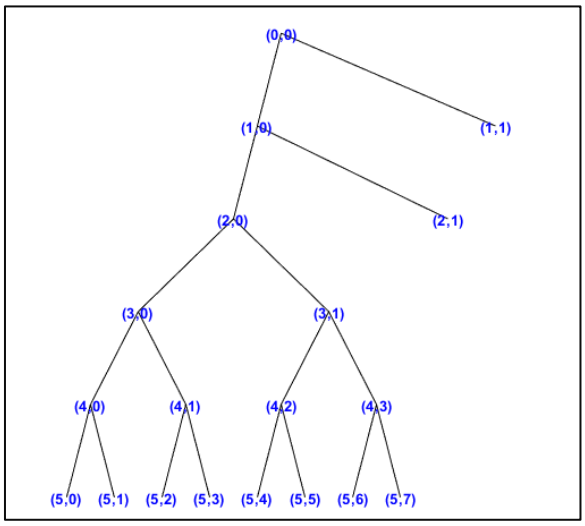


Fig 4: Pruned Binary tree structure depicting the signal decomposition

A 5-level binary tree decomposition approach is implemented to separate the various sub-bands in the signal, in order to distinctly capture the 450Hz signal component being analyzed. The Symlet 8(Sym8) wavelet is used in islanding detection because it has a good balance between time and frequency localization, making it suitable for detecting small changes in the signal overtime during islanding detection procedure. Additionally, the Sym8 wavelet has a higher number of vanishing moments compared to other commonly used wavelets, which allows for better suppression of noise and better signal reconstruction after wavelet packet decomposition, compensating the shortcoming of the Fuzzy inferencing system methodology [26]. I_{ro1} current waveform when under different cases of system operation, i.e., system is operating under normal operation, when DER_2 is disconnected, when DER_3 is disconnected and when both DERs are disconnected are as shown in Fig 5, Fig 6, Fig 7, and Fig 8 respectively. Fig 6, Fig 7, and Fig 8, evidently depict the disconnection of DER_1, DER_2, and DER_3 respectively from the system t = 833 min respectively.

The I_{ro1} 9th harmonic sub-band under normal operation is shown in Fig 9. Fig 9 has a magnitude of zero, because, under normal condition of operation, all the 9th harmonic components add up to zero, indicating absence of islanding. The 9th harmonic sub-band of I_{ro1} when DER_2 is islanded is shown in Fig 10. The 9th harmonic sub-band current signature of I_{ro1} when DER_3 is islanded is shown in Fig 11. The 9th harmonic sub-band current signature in the above two cases is different, owing to the harmonic pilots' injection method proposed in Fig 3. The phase displacement introduced during current pilot injection leads to the acquisition of distinct and individual signatures within the line current waveforms at different nodes. The 9th harmonic sub-band current signature of I_{ro1} when both DER_2 and DER_3 are islanded shown in Fig 12, is different to that shown in Fig 8, Fig 9 and Fig 10. The same analysis can be applied for the detection of islanding of upstream DERs by detecting the V_{pcc} at the respective PCC of the DERs lying downstream of the disconnected DER. Since this method does not require any communication infrastructure, the ratio of accuracy of detection to the cost of implementation is very high.

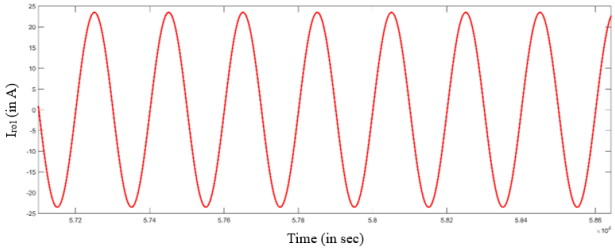


Fig 5: I_{ro1} waveform under normal condition of operation

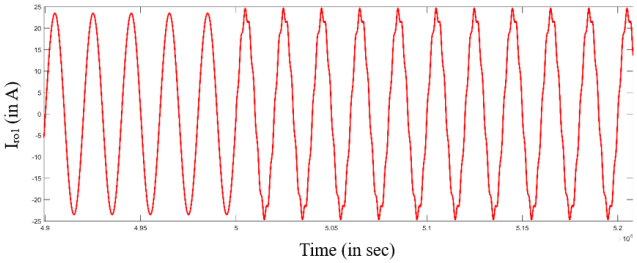


Fig 6: I_{ro1} waveform when DG_2 is disconnected

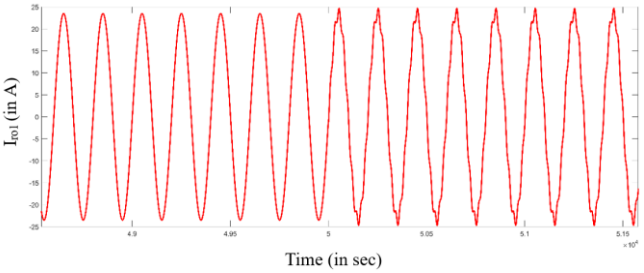


Fig 7: I_{ro1} waveform when DG_3 is disconnected

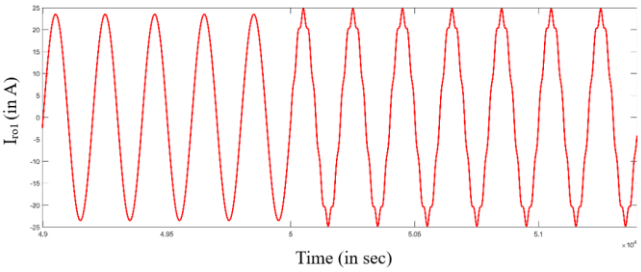


Fig 8: I_{ro1} waveform when both DG_2 and DG_3 are disconnected

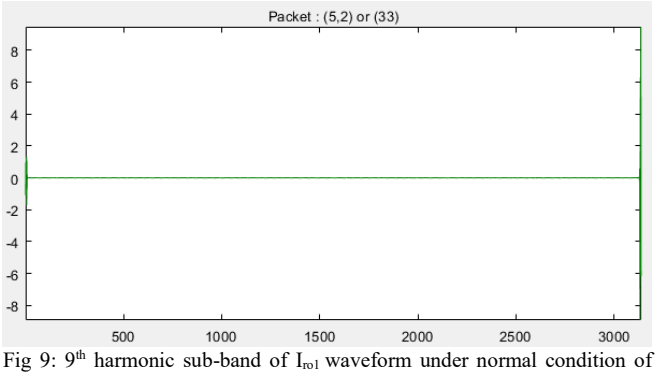


Fig 9: 9th harmonic sub-band of I_{ro1} waveform under normal condition of operation

The WPT divides the power signal into various frequency sub bands through multiple decomposition levels. When an islanding event of grid occurs, distinct frequency components or spectral patterns can arise because of the disconnection of the utility grid. WPT can effectively detect

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:16:10 UTC from IEEE Xplore. Restrictions apply.

and analyze these frequency variations, offering insights into the modified spectral characteristics of the power signal.

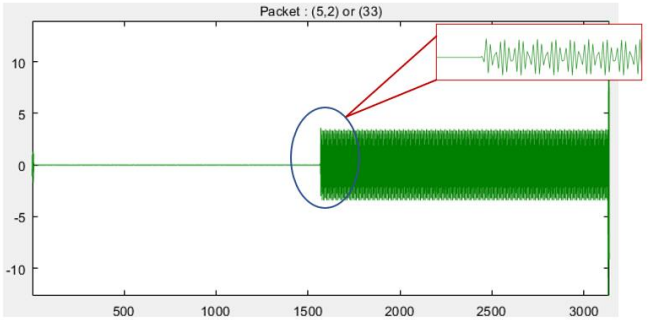


Fig 10: 9th harmonic sub-band of I_{ro1} waveform when DG_2 is disconnected

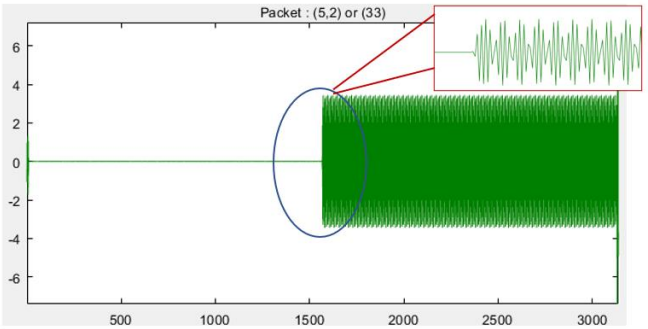


Fig 11: 9th harmonic sub-band of I_{ro1} waveform when DG_3 is disconnected

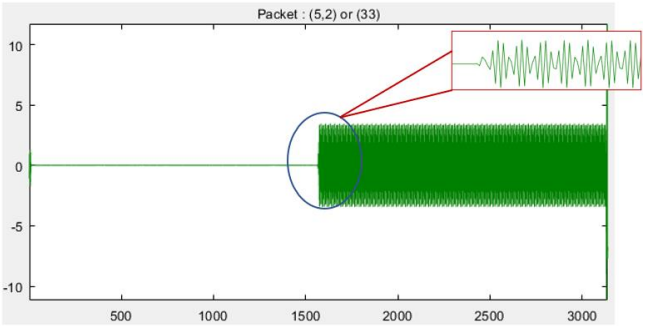


Fig 12: 9th harmonic sub-band of I_{ro1} waveform when both DG_2 and DG_3 are disconnected

IV. CONCLUSION

The proposed WPT based grid islanding cum DG outage detection technique incorporating Wavelet Packet Transform has been verified using MATLAB/Simulink based results. This method addresses the issue of sensitivity to noise applicable to frequency-based islanding detection method. The proposed method is also simple and does not become complex when extending this technique to large number of DGs. Thus, simulation results obtained support the theoretical explanation of efficacy of current and voltage signatures method of islanding detection method. The suggested WPT method of islanding detection aids in accurate discrimination between different disturbances owing to the finer resolution it offers in both time and frequency domains. This makes it robust to load changes or the presence of harmonic disturbances. Further, this technique can be implemented in collaboration with other islanding detection techniques, to seek better performance.

REFERENCES

[1] A. F. Robert and R. C. Dugan, “Distribution system analysis and the future smart grid,” IEEE Trans. Ind. Appl., vol. 47, no. 6, pp. 2343–2350, Dec. 2011.

[2] W. Xiaoyu and W. Freitas, “Impact of positive-feedback anti-islanding methods on small-signal stability of inverter-based distributed generation,” IEEE Trans. Energy Convers., vol. 23, no. 3, pp. 923–931, Sep. 2008.

[3] S. Baloch, S. S. Samsani and M. S. Muhammad, "Fault Protection in Microgrid Using Wavelet Multiresolution Analysis and Data Mining," in IEEE Access, vol. 9, pp. 86382-86391, 2021, doi: 10.1109/ACCESS.2021.3088900.

[4] V. Verma and G. G. Talpur, "Decentralized Master-Slave operation of microgrid using current controlled distributed generation sources", *2012 IEEE International Conference on Power Electronics Drives and Energy Systems (PEDES)*, pp. 1-6, 2012.

[5] IEEE Standard for Interconnection and Interoperability of Distributed En[1]ergy Resources with Associated Electric Power Systems Interfaces, IEEE Standard 1547-2018 (Revision IEEE Standard 1547-2003), pp. 1–138, 2018.

[6] “UL standard for inverters, converters, controllers and interconnection system equipment for use with distributed energy resources,” UL 1741, Ed. 3, Sep. 2021.

[7] H. H. Zeineldin, T. Abdel-Galil, E. F. El-Saadany, M. M. A. Salama, Islanding detection of grid connected distributed generators using TLS-ESPRIT, Electric Power Systems Research, 77 (2) (2007)155- 162.

[8] M. Valentini, S. Munk-Nielsen, F. Valderrey Sanchez, U. Martinez De Estibariz, “A new passive islanding detection method for grid-connected PV inverters”, in Proc. of the International Symposium on Power Electronics, Electrical Drives, Automation and Motion, 2008, pp. 223-228.

[9] H. Zeineldin and J. Kirtley, "A simple technique for islanding detection with negligible non-detection zone," 2009 IEEE Power & Energy Society General Meeting, Calgary, AB, Canada, 2009, pp. 1-1, doi: 10.1109/PES.2009.5275939.

[10] M. Liserre, A. Pigazo, A. Dell'Aquila and V. M. Moreno, "An Anti-Islanding Method for Single-Phase Inverters Based on a Grid Voltage Sensorless Control," in IEEE Transactions on Industrial Electronics, vol. 53, no. 5, pp. 1418-1426, Oct. 2006, doi: 10.1109/TIE.2006.882003.

[11] J.A. Laghari, H. Mokhlis, M. Karimi, A.H.A. Bakar, Hasmaini Mohamad, “Computational Intelligence based techniques for islanding detection of distributed generation in distribution network: A review”, Energy Conversion and Management, Volume 88, 2014, Pages 139-152, ISSN 0196-8904, https://doi.org/10.1016/j.enconman.2014.08.024.

[12] Y. Bansal and R. Sodhi, "A Statistical Features Based Generic Passive Islanding Detection Scheme for IIDGs System," in IEEE Transactions on Power Delivery, vol. 37, no. 4, pp. 3176-3188, Aug. 2022, doi: 10.1109/TPWRD.2021.3124986.

[13] Byunggyu Yu, Mikihiko Matsui, Youngseok Jung and Gwonjong Yu, “A combined active anti[1]islanding method for photovoltaic systems”, Renewable Energy, vol. 33, no. 5, pp. 979-985, May 2008.

[14] L. A. C. Lopes and Y. Zhang, "Islanding Detection Assessment of Multi-Inverter Systems With Active Frequency Drifting Methods," in IEEE Transactions on Power Delivery, vol. 23, no. 1, pp. 480-486, Jan. 2008, doi: 10.1109/TPWRD.2007.911188.

[15] H. Zheng, Z. Liu, R. An, J. Liu, T. Wu and Z. Lin, "An Islanding Detection Method Using Synchronized Small-AC-Signal Injection for Grid-Forming Inverters in Microgrids," in IEEE Transactions on Power Electronics, vol. 38, no. 5, pp. 5816-5831, May 2023, doi: 10.1109/TPEL.2023.3237569.

[16] T. Shi et al., "Detecting Speed Improvement and System Stability Enhancement for DC Microgrids Islanding Detection Based on Impedance Characteristic Analysis," in IEEE Transactions on Power Electronics, vol. 38, no. 3, pp. 3785-3802, March 2023, doi: 10.1109/TPEL.2022.3221741.

[17] H. R. Baghaee, D. Mlakić, S. Nikolovski and T. Dragicević, "Support Vector Machine-Based Islanding and Grid Fault Detection in Active Distribution Networks," in IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 8, no. 3, pp. 2385-2403, Sept. 2020, doi: 10.1109/JESTPE.2019.2916621.

[18] X. Chen, Y. Li and P. Crossley, "A Novel Hybrid Islanding Detection Method for Grid-Connected Microgrids With Multiple Inverter-Based

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:16:10 UTC from IEEE Xplore. Restrictions apply.

Distributed Generators Based on Adaptive Reactive Power Disturbance and Passive Criteria," in IEEE Transactions on Power Electronics, vol. 34, no. 9, pp. 9342-9356, Sept. 2019, doi: 10.1109/TPEL.2018.2886930.

[19] P. Kumar, V. Kumar and B. Tyagi, "Sequence-Based Hybrid Technique for Islanding Detection for Microgrid With RES," in IEEE Transactions on Industry Applications, vol. 58, no. 1, pp. 185-200, Jan.-Feb. 2022, doi: 10.1109/TIA.2021.3130214.

[20] V. R. Reddy and E. S. Sreeraj, "A Hybrid Islanding Detection Method for One Cycle Controlled PV Inverter System," in IEEE Journal of Emerging and Selected Topics in Industrial Electronics, vol. 3, no. 3, pp. 777-787, July 2022, doi: 10.1109/JESTIE.2021.3099395.

[21] K. Dhibi et al., "A Hybrid Fault Detection and Diagnosis of Grid-Tied PV Systems: Enhanced Random Forest Classifier Using Data Reduction and Interval-Valued Representation," in IEEE Access, vol. 9, pp. 64267-64277, 2021, doi: 10.1109/ACCESS.2021.3074784.

[22] M. Seyedi, S. A. Taher, B. Ganji and J. Guerrero, "A Hybrid Islanding Detection Method Based on the Rates of Changes in Voltage and Active Power for the Multi-Inverter Systems," in IEEE Transactions on Smart Grid, vol. 12, no. 4, pp. 2800-2811, July 2021, doi: 10.1109/TSG.2021.3061567.

[23] S. Murugesan and V. Murali, "Hybrid Analyzing Technique Based Active Islanding Detection for Multiple DGs," in IEEE Transactions on Industrial Informatics, vol. 15, no. 3, pp. 1311-1320, March 2019, doi: 10.1109/TII.2018.2846025.

[24] H. T. Do, X. Zhang, N. V. Nguyen, S. S. Li and T. T. -T. Chu, "Passive-Islanding Detection Method Using the Wavelet Packet Transform in Grid-Connected Photovoltaic Systems," in IEEE Transactions on Power Electronics, vol. 31, no. 10, pp. 6955-6967, Oct. 2016, doi: 10.1109/TPEL.2015.2506464.

[25] V. Verma and S. Jain, "Active islanding detection method for grid and DG failure in a community Microgrid application," 2016 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), Trivandrum, India, 2016, pp. 1-5, doi: 10.1109/PEDES.2016.7914524.

[26] S. Swetha and V. Verma, "Autonomous Active Grid islanding and DGs outage detection in a $\mu \mathrm{G}$ utilizing Goertzel Algorithm in collaboration with Fuzzy Inferencing System," 2023 IEEE IAS Global Conference on Renewable Energy and Hydrogen Technologies (GlobConHT), Male, Maldives, 2023, pp. 1-6, doi: 10.1109/GlobConHT56829.2023.10087768.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:16:10 UTC from IEEE Xplore. Restrictions apply.

Cluster Computing
https://doi.org/10.1007/s10586-024-04266-0

Deployment and performance monitoring of docker based federated learning framework for software defect prediction

Ruchika Malhotra[1] · Anjali Bansal[1] · Marouane Kessentini[2]

Received: 1 September 2023 / Revised: 30 December 2023 / Accepted: 3 January 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
There are significant challenges in machine learning models due to information security and data privacy issues. In traditional machine learning approaches, the data used to train the centralized model may be sensitive leading to privacy issues. Federated learning overcomes these issues by following the concept of bringing the code to the data instead of data to the code. In today's era of digital transformation, researchers and industries are focusing on cloud-based technologies. New approaches such as microservice models, docker, kubernetes, containers, and virtual machines are being used by IT industries to develop and deploy applications. We make use of software containers to deploy federated learning models and then we identify security vulnerabilities and resource usage in federated learning model. Flask, docker, cAdvisor, prometheus, and grafana are integrated to monitor the resource usage of the federated learning model in terms of CPU utilization, memory utilization, network traffic usage, and disk usage. The generated federated learning model is deployed on docker, and security vulnerabilities and performance are analyzed using open-source tools. The result shows that the generated model is secure and performs efficiently.

Keywords Federated learning · Flask · Docker · Software defect prediction · Monitoring tools · Security vulnerabilities

1 Introduction

Software development life cycle consists of software testing which is time and resource-consuming. The testing procedure aims to deliver software that is completely error-free and meets the needs of all stakeholders. Testing adds to the overall project budget because it is an expensive process. The effectiveness and quality of the software improve when faults are correctly predicted at an early stage [1]. Finding software defects is a necessary but expensive process. Many approaches such as machine learning (ML) techniques, ensemble techniques, deep learning (DL) techniques, and data mining techniques have been introduced to predict the defects. Giray et al. [2] performed a literature survey on 102 studies to analyze the use of DL techniques in software defect prediction. They concluded that a large number of studies used supervised DL techniques and CNN is the most used DL technique in software defect prediction.

Although ML and DL approaches have proven their efficiency in all areas but these approaches do not perform well on decentralized data. Federated learning (FL) is gaining significant attention nowadays as it works well on distributed and decentralized data and this approach also maintains privacy of the data. Data security and data privacy is an essential issue that needs to be resolved. In FL, various clients work together to solve ML issues while being coordinated by a central server. The idea of FL has appeared frequently in various forms throughout the history of ML and computer science. For example, collaborative deep learning, privacy-preserving machine learning, federated optimization, distributed deep learning, and privacy-preserving data analytics. FL is a decentralized ML

✉ Anjali Bansal
anjalibansal791@gmail.com

Ruchika Malhotra
ruchikamalhotra@dtu.ac.in

Marouane Kessentini
kessentini@oakland.edu

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

[2] Computer Science Department, Oakland University, Rochester, MI, USA

Published online: 20 February 2024

approach in which the model moves to data instead of data moves to the model.

Building a DL model that works on one system is not a difficult task. However, it is more challenging to generate a DL model that can run on any system around the world without downloading any dependencies or libraries. Organizations are moving towards DevOps as well as continuous integration and continuous deployment practices for software/application development. Virtualization enables continuous software development and deployment on virtual hardware while container-based virtualization (containerization) enables software applications to operate on a single operating system in a virtual run-time environment. Nowadays, IT industries make use of containers to deploy the applications/models. The use of containers for model deployment provides modularity, portability, and reusability. This paper introduced a way to deploy a deep learning model in microservice architecture using docker to use it anywhere around the world.

In this paper, a defect prediction model is generated using the FL approach. Although DL frameworks are widely used in software vulnerability detection and in solving many real-world problems, these frameworks may have some vulnerabilities that can harm the security of learning models. A vulnerability refers to any flaws and weaknesses in the system due to which an attacker can gain control of the system. These vulnerabilities can occur due to two reasons: either there are some errors in code or there is some defect in the design of the system.

In recent years, researchers are focusing on the security of DL models. The main objective of this study is to examine the security vulnerabilities and performance of the federated transfer learning model in terms of CPU utilization, memory utilization, network traffic usage, and disk usage.

There exist several container scanning tools such as AppArmor, Anchore, Microscanner, Clair, Cilium, JFrog Xray, Dagda, Aqua Trivy, Qualys, Snyk, etc. which are used to detect security vulnerabilities in containers. Containerization has gained significant traction in modern software development, enabling enhanced scalability and portability of applications. However, monitoring the performance of containerized applications presents new challenges. There exist various benchmark tools such as ycruncher [3], linpack [4], geekbench [5], sysbench [6], stream [7], bonniee + + [8], fio [9], netperf [10], iperf [11], etc. and monitoring tools such as docker stats [12], sysdig [13], datadog [14], sematext [15], cAdvisor, Prometheus, Grafana, etc. to analyze the performance of containers. In this study, we employ Prometheus, Grafana, and cAdvisor to tackle these challenges and gain deeper insights into the containerized application's behavior. We examine the performance of the federated transfer learning-based DL model in terms of CPU utilization, memory utilization, network traffic usage, and disk usage.

This paper mainly focuses on the following research questions:

RQ1: What is the difference between the performances of the FL model having CNN as the base model and the FL model having ANN as the base model?
RQ2: How vulnerable is the FL model?
RQ3: How does the FL model perform in terms of CPU utilization, memory utilization, network traffic usage, and disk usage?

RQ1 examines the prediction performance of FL with CNN as base model and FL with ANN as base model. RQ2 is related to analyzing the security vulnerabilities in the generated FL model. In this study, anchore, aqua trivy, jFrog xray, docker scout, and docker scan tools are used for security vulnerability analysis. RQ3 is related to analyzing the resource usage by generated FL model. In this study, locust, docker stats, cAdvisor, sematext, prometheus, and grafana tools are used to evaluate the resource utilization.

The organization of the rest of the paper is as follows: The background knowledge about the FL approach, docker container, and software defect prediction is described in Sect. 2. The existing studies related to FL, software defect prediction, and containers are described in Sect. 3. The methodology used to evaluate the security and performance of the FL model is described in Sect. 4. Section 5 describes the results based on experimental analysis and visualization graphs. Conclusion and future scope is described in Sect. 6.

2 Background

In this section, background knowledge of docker container, FL, and software defect prediction concepts is described.

2.1 Docker

Docker is an open-source containerization platform for deploying and executing applications/services. Docker has become the industry standard for software development and deployment because it makes application packaging and delivery easier while enhancing reproducibility and consistency. Docker containers are self-contained, lightweight, executable, and portable packages that consist of everything needed to run a piece of software, including the code, runtime, libraries, and system tools. Containers are based on containerization technology, which allows applications to be isolated from their environment and ensures that they run consistently across various computing environments. Docker containers differ from the virtual machine as container virtualizes the operating system but

virtual machine follows hypervisor-based virtualization and it virtualizes hardware. Figure 1 shows the difference between containers and virtual machines. In hypervisor-based virtualization, each application requires a separate operating system but in containerization, applications share an operating system, so the container-based virtualization technique is lightweight and deployment can be smaller in size. Containers are replacing virtual machines as containers provide a shared host operating system, fast deployment, quick launch time, portability, and scalability.

2.2 Federated learning

Centralized learning and FL are two different approaches to build ML models. Centralized learning is the traditional approach to build ML models where data is collected and stored at the server and the ML model is trained at the server and then deployed at the server to make predictions. This leads to privacy threats to user's data stored on the server. So, Google introduced a new approach known as FL to provide privacy to user's local data. FL is a decentralized ML approach where the data is distributed across many client devices or nodes and the model is trained on the client device itself.

Instead of sending all the data to a central server for training, the model is sent to the data sources, and the model parameters are updated locally. Only the updated model parameters are then sent back to the central server, where they are aggregated to enhance the global model. This ensures that data in each client device adheres to privacy policies. In other terms, we can say that in the traditional ML approach, data moves to the model but in FL, the model moves to data. Figure 2 shows the difference between centralized learning and FL. FL is particularly useful in situations where data sharing raises privacy concerns and the data is too large or sensitive to be centralized. Zhang et al. [16] and Li et al. [17] reviewed all the existing studies related to FL. They discussed the challenges, future directions, and applications of FL. It has been used in wireless communication [18–21], healthcare systems [22–25], and mobile devices for recommending the services [26–28].

2.3 Software defect prediction

Defect prediction and proneness in the software are always considered a major problem in the software industry and for software engineers. Software defect prediction is a field of software engineering that aims to detect and anticipate software defects or bugs before they arise during software development. By identifying parts of the code that are more likely to have errors, the goal is to proactively manage and enhance software quality and enable developers to allocate resources for testing and debugging more effectively. In classical methods, previous experience with defective or non-defective software is required while detecting the software defect in a software application [1]. While software defect prediction can be a useful tool, it should be used in conjunction with other software testing and quality assurance procedures as it cannot guarantee the discovery of all faults. Additionally, the success of defect prediction models depends on the availability and quality of historical

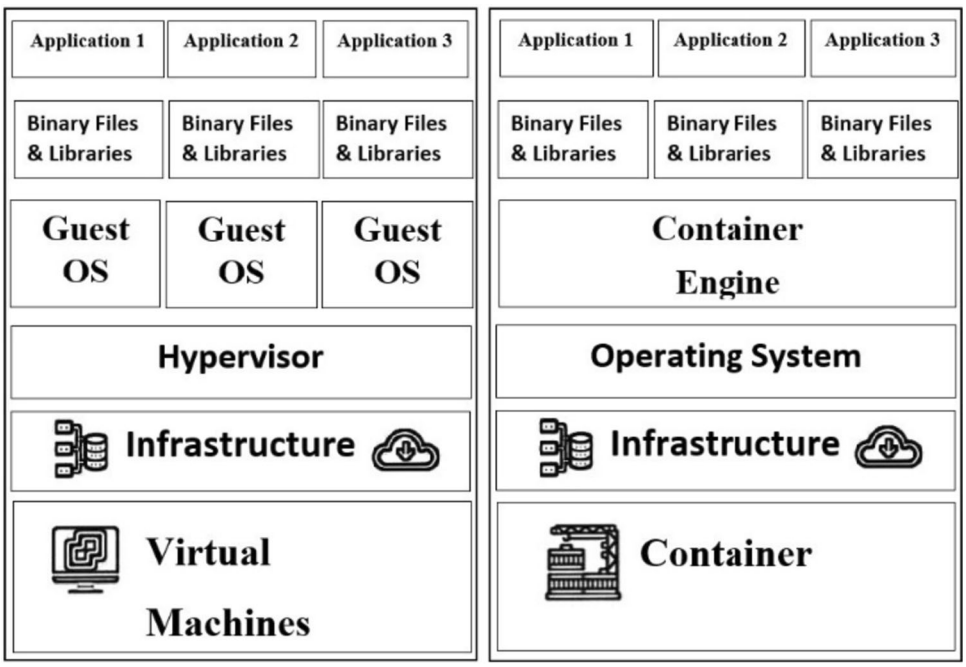


Fig. 1 Container versus virtual machine

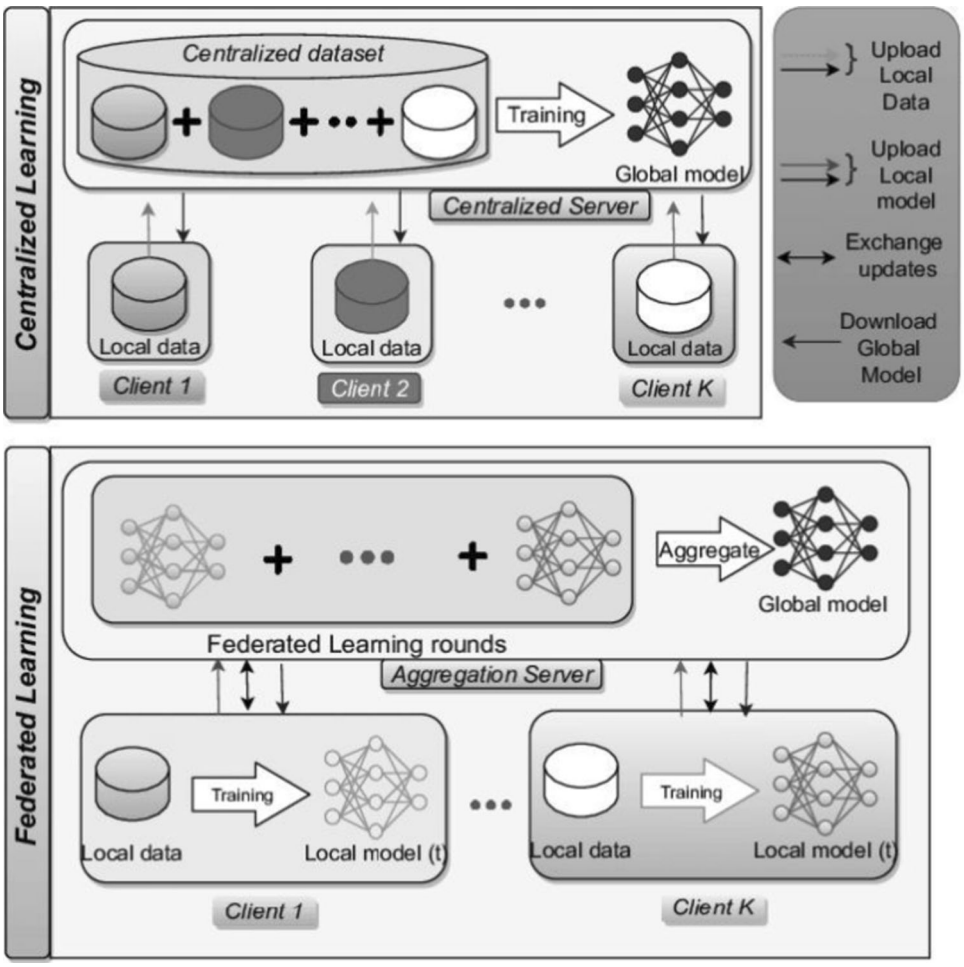


Fig. 2 Centralized versus federated learning

data and the relevance of selected features. Defects in the software applications have been predicted by using different classification techniques by different researchers as of now but the results of the methods changed with the dataset, so they lack the property of robustness for defect prediction in an unknown software application. The field of software defect prediction has witnessed a lot of literature reviews published. The majority of them employ static code metrics, and a few studies demonstrate the connection between fault proneness and process metrics.

In summary, the FL approach is used to build the software defect prediction model in order to maintain the privacy and security of the user's data. Docker is used to deploy the generated model. Docker is widely adopted in the software development and deployment field as the industry standard because it makes the packaging and delivery of applications easier and improves reproducibility and consistency.

3 Related work

Many models have been introduced to predict defects in software using ML and DL approaches. This section provides a brief summary of previous studies on software defect prediction and FL. Furthermore, this section also mentions previous studies and existing tools related to security and performance analysis of the model.

3.1 Software defect prediction

In the last few decades, there has been an increase in research work in the field of defect prediction as defects play a significant role in the software development process. Researchers used product metrics and process metrics as measures of software characteristics. Komalasari and Candra [29] used a combined model of product, process, and profile metrics to build a classification model. They used Naïve Bayes, Logistic Regression, and Random Forest techniques to predict the defects. The result shows that a combination of these three metrics gives effective results in the Random Forest model. Choudhary et al. [30]

proposed new change metrics and analyzed the effect of existing and proposed change metrics in software defect prediction on the eclipse dataset. They used decision tree, KNN, and random forest as classification techniques and precision, recall, and F-measure as performance evaluation metrics. They analyzed that the combined model of change and code metrics outperforms the model that has an individual set of metrics on the eclipse dataset. Rhmann et al. [31] analyzed the performance of hybrid search-based algorithms for software defect prediction using change metrics. They used the android dataset and recall and precision as performance evaluation measures. The result shows that hybrid search-based algorithms perform better as compared to machine learning techniques.

3.2 Federated learning

Nowadays, researchers are focusing on data security and privacy-preserving approaches. Zhou et al. [32] proposed a novel approach based on FL and knowledge distillation for bearing defect prediction. They performed experiments on public datasets of Paderborn University and Case Western Reserve University. The result shows that the proposed approach gives higher accuracy as compared to existing approaches. Wang et al. [33] proposed a federated transfer learning approach (FTLKD) to predict the defects in the software modules where data is heterogeneously collected from various sources. They used NASA, AEEEM, and SOFTLAB datasets. They used knowledge distillation to provide the communication between clients. Experimental result of this study shows that the proposed approach outperforms existing techniques in terms of AUC, accuracy, and Gmean values. Wang et al. [34] proposed a federated learning approach (FRLGC) which uses deep reinforcement learning to train the private models on public data to predict the defects in software modules. They used Relink, NASA, and AEEEM datasets. The result shows that the proposed approach outperforms existing heterogeneous software defect prediction techniques in terms of AUC and Gmean values. Wang et al. [35] proposed a federated sparse compression algorithm to reduce the cost of communication. They performed experiments on Relink, NASA, and AEEEM datasets. They used differential privacy and capsule networks. The result depicts that it performs better in terms of AUC and Gmean values with less communication bits. Nandi et al. [36] introduced DFL (Docker-enabled Federated Learning) framework for processing data streams on heterogeneous clients. They tested the effectiveness of DFL on real-time emotion classification problems. They used accuracy and F1 measure to evaluate the performance of DFL. The result demonstrates that the performance of DFL is good and it can handle multiple clients running simultaneously.

3.3 Security and performance analysis

He et al. [37] conducted an extensive and comprehensive survey to investigate various attacks on deep learning models. They summarized four types of attacks in their study: model extraction attack, model inversion attack, poisoning attack, and adversial attack. Xiao et al. [38] performed an empirical study to detect the set of vulnerabilities and their corresponding risks in three deep learning frameworks (Caffe, Torch, and TensorFlow). The types of risks include evasion of detection, denial of service, and system compromise. Javed and Toor [39] in their study investigated the quality of various container scanning tools based on coverage and accuracy. Morabito et al. [40] analyzed the performance of containers using various benchmark tools such as ycruncher, linpack for CPU usgae, stream for memory usage, bonniee + + for disk usage, and netperf for network usage. Kozhirbayeu and Sinnott [41] investigated CPU utilization with the help of linpack, geekbench, and ycruncher, memory utilization with the help of stream, disk utilization with the help of bonniee + + and random write speed, and network utilization with the help of netperf and iperf. They concluded that there is no overhead on CPU utilization and memory utilization. Matti Holopainen [42] introduced a monitoring setup to monitor the resource usage (CPU usage, memory usage, and network bandwidth usage) of the container environment using Prometheus, and Grafana. cAdvisor and node exporter are used as data collection tools in this study. Tiia Leppänen [43] explored and compared Prometheus and Grafana with Zabbix and Libre NMS monitoring tools. Node exporter and SNMP exporter are used in this study for data collection. The advantages of Prometheus and Grafana are also discussed in this study.

This brief literature provides a discussion about software defect prediction models using ML, DL and FL approaches. We noticed that most of the proposed models used only the dataset that consist of only static code metrics. In this work, we introduced FL approach for the dataset containing both static code metrics and process metrics. This section also provides a brief discussion about approaches and tools used to evaluate the security and resource usage of docker containers.

4 Experimental method

In this study, the FL model having CNN and ANN as base models is proposed for predicting software defects. We analyze the security vulnerabilities in FL model with the help of open-source tools. CPU utilization, memory utilization, network traffic usage, and disk usage are also

analyzed with the help of monitoring tools. The experiment is performed on an open-source dataset. This section describes the steps followed to perform the experiments including dataset description, data preprocessing, deep learning models used, performance evaluation measure used, tools used for security analysis, and monitoring tools used to evaluate the performance. Figure 3 describes various steps followed to perform the experiment.

4.1 Dataset used

The initial goal of the study was to use publicly available dataset to promote replicability of this work. The dataset used in this study is collected from the publicly available metric repository which is maintained by Jureczko and Madeyski [44] which consists of both static code metrics and process metrics.

Static code metrics talk about the design complexity and size of the source code, while the complexity of the software development process is described by process metrics. The following static code metrics are used in this study: weighted methods per class (WMC), depth of inheritance tree (DIT), number of children (NOC), coupling between object (CBO), response for a class (RFC), lack of cohesion in methods (LCOM), afferent coupling (Ca), efferent coupling (Ce), lines of code (LOC), number of public methods (NPM), data access metric (DAM), measure of aggregation (MOA), cohesion among methods of class (CAM), Inheritance coupling (IC), coupling between methods (CBM), average method complexity (AMC), and McCabe's cyclomatic complexity (CC). The definition of the aforementioned metrics is given in [45]. The process metrics used in this study are as follows: number of revisions (NR), number of distinct committers (NDC), number of modified lines (NML), and number of defects in previous versions (NDPV). The definition of the aforementioned metrics is given in [46].

We choose 4 java-based projects, which provides the consistency in terms of the programming languages used. In this study, 8 versions of 4 projects have been used for experiment. Basically, the original datasets had labels showing the number of defects. This label is replaced into a simpler system: if there is no defect, it is labeled as non-defective or 0, and if there are one or more defects, it is labeled as defective or 1. Table 1 lists the datasets that are used in this study which shows the project name with their version (column 1). Then it shows number of instances, number of faulty instances, and number of non faulty

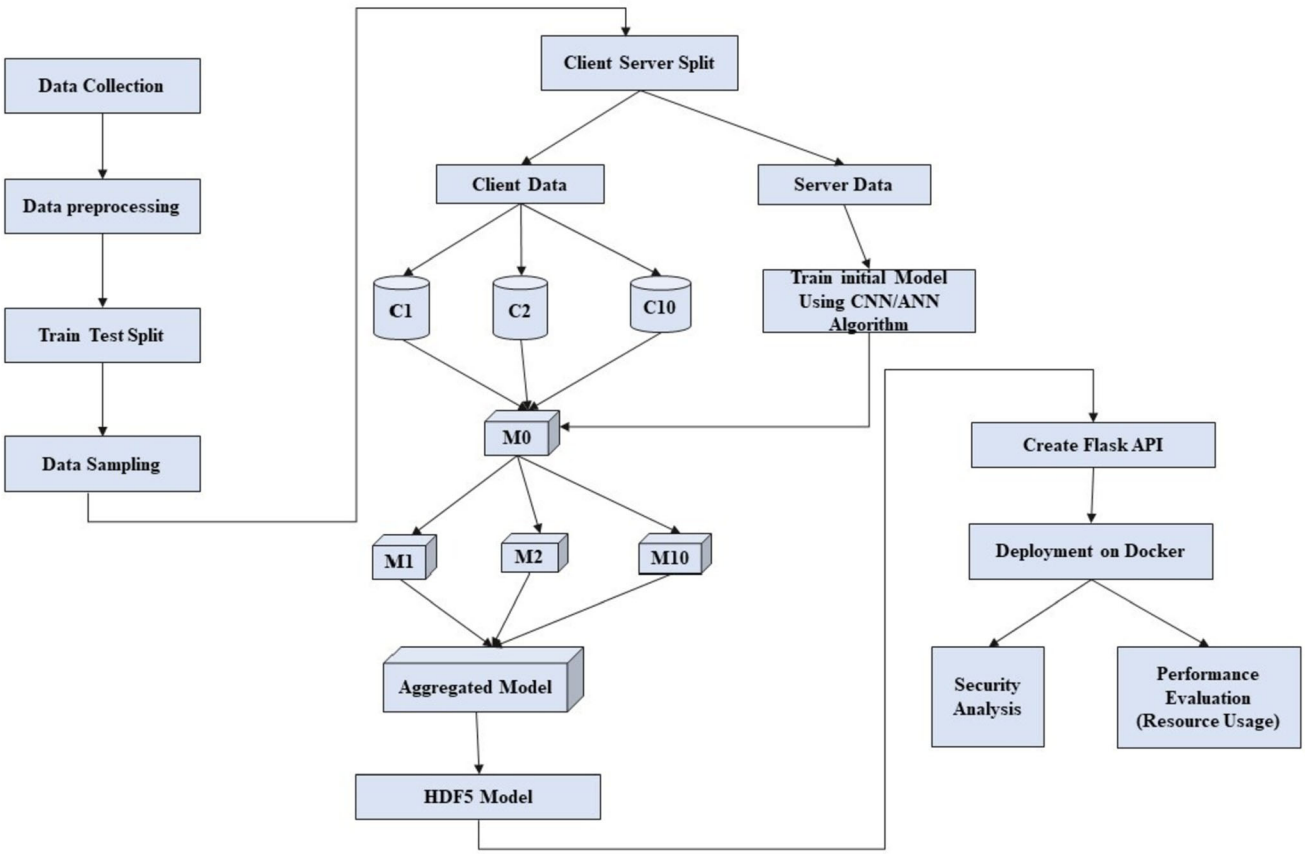


Fig. 3 Architecture of proposed methodology

Table 1 Dataset used

Project	Number of instances	Faulty instances	Non-faulty instances	Description
Ant 1.6	524	92	432	Java based build tool
Ant 1.7	1065	166	899	
jEdit 4.1	644	79	565	Java based cross platform text editor
Xalan 2.5.0	945	387	558	XSLT processor
Xalan 2.6.0	1170	411	759	
Xerces 1.2.0	515	71	444	XML parser
Xerces 1.3.0	545	69	476	
Xerces 1.4.4	671	437	234	

instances (column 2–4). Finally, it shows the description of the project used (column 5).

4.2 Dataset preprocessing

Datasets obtained from publicly available repositories may contain missing values that can negatively impact the performance of the created model. Hence preprocessing is done on datasets to avoid this type of problems. This can include eliminating data points with missing values if they are extremely small in count, or replacing missing values with mean value or median value or with some user defined constant.

There is an imbalance class problem in the datasets chosen. The imbalanced data identification is done by taking the ratio of a number of Non-Defective and Defective Classes in the train dataset since the ratio should be nearly equal to 1: 1 for balanced data. So, random oversampling of the data is done to balance it. In random oversampling, the weightage of the class which is low in the count is increased by repetition of that class's data points randomly. The defective class is very low in count in all the datasets used, so it should be given more weightage by repetition to balance the dataset. If the ratio of non-defective to defective class is X: 1 then defective class's data point will be repeated X times to balance the data and to avoid biasness of the majority class. Figure 4 shows the data before oversampling and after oversampling. Random Over Sampling of the minority class is done (only during the training of the model).

4.3 Algorithm used

In this study, FL approach for predicting software defects is used. We use deep learning models CNN and ANN as base models. From Fig. 3, firstly base deep learning model CNN/ANN is trained using server data to build the global model (M0). The global model is sent to each client and clients train the global model using their local data on their device to build a local model (M1, M2…, M10). All the local model weights are aggregated using FedAvg (Federated Averaging) method to build a global model and then the global model is sent back to the server. The performance of ANN based FL model and CNN based FL model is compared. We use Adam optimizer with relu activation function for all the layers except the output layer. As the dependent variable has binary classes 0 or 1, so sigmoid activation function is used in the output layer. Table 2 shows hyperparameter details of the proposed framework for predicting software defects.

4.3.1 Artificial neural network

An artificial neural network (ANN) is a type of DL algorithm inspired by the structure and function of the human brain and is used in various fields such as speech recognition, computer vision, and finance. ANN is composed of an input layer, one or two hidden layers, and an output layer. The input layer consists of independent variables and the output layer acts as the dependent variable. Here, we use 2 hidden layers with a dropout rate of 0.05.

4.3.2 Convolutional neural network

Convolutional neural network (CNN) is a type of DL neural network that consists of multiple hidden layers. CNNs are also known as CovNets. In CNN, the input data is of image type and CNN makes use of convolutional kernels and filters to extract the features of the image such as edges, textures, and corners. The extracted features are then passed through multiple layers of the network to make a prediction or classification. CNNs are widely used in image recognition, natural language processing, medical imaging, and object detection.

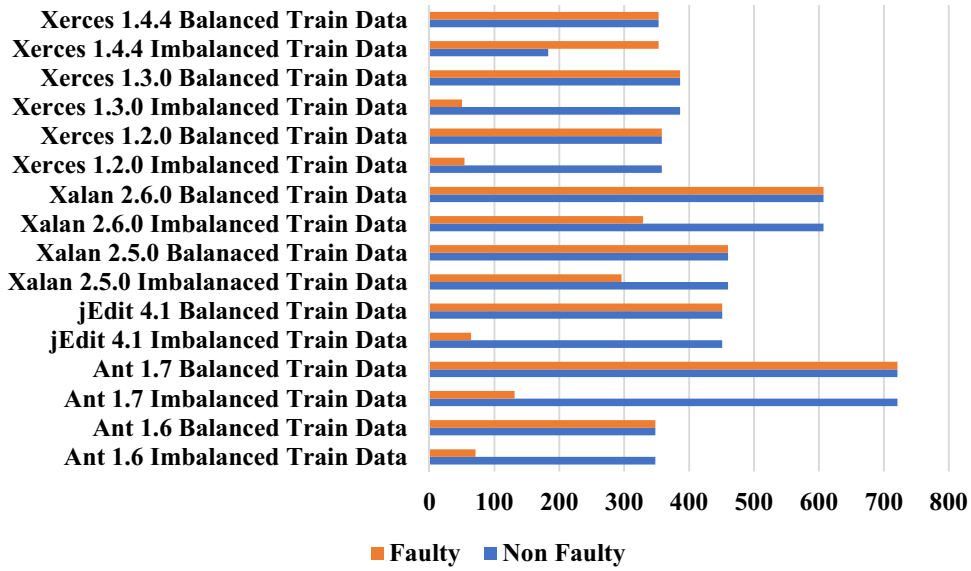


Fig. 4 Before oversampling and after oversampling

Table 2 Hyperparameter details

Components of the proposed framework	Hyperparameters
Number of iterations	20
Number of clients	10
Verbose	0
Epochs	100
Batch size	10
Activation function	ReLu, Sigmoid (At last level)
Optimizer	Adam
Loss function	Binary cross entropy
Number of convolution 2D layers	3
Regularization method	Dropout

4.4 Performance evaluation measure

Area under ROC (AUC) and Matthew's correlation coefficient (MCC) are used to evaluate the prediction performance of federated learning models. AUC is a summarization of the ROC curve [47]. It plots the curve between two parameters: TPR and FPR. The value of AUC lies between 0 and 1. MCC metric indicates the correlation coefficient between predicted and actual observation [48]. Equation 1 represents the formula for calculating the MCC value. The value of MCC lies between -1 and 1. We can find values of TPR, FPR, true positive (TP), true negative (TN), false positive (FP), and false negative (FN) with the help of a confusion matrix.

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (1)$$

4.5 Tools used for security analysis

There can be some security issues in federated learning models as some packages can be outdated or may have some vulnerabilities. To select container scanning tools, we followed a set of criteria. We choose widely-used, freely-available tools that have been utilized by practitioners or researchers. Free tools are considered due to the limited capabilities of trial versions of paid products. Here, Anchore, Aqua Trivy, JFrog Xray, Docker Scout, and Docker scan are used to detect security vulnerabilities in the docker image which consists of all the files, libraries, and packages needed in building the federated learning model for software defect prediction. We have installed extensions of tools used in this study on the docker desktop and then the created docker image is scanned to detect security vulnerabilities. Table 3 lists the features of container scanning tools used [49].

(a) Anchore: Anchore is an open-source tool that examines every layer of docker image to identify

Table 3 Container scanning tools used

Tool name	OS packages	Non OS packages	Database information
Anchore	Yes	Yes	No
Aqua trivy	Yes	Yes	Yes
jFrog xray	Yes	Yes	Yes
Docker scan	Yes	Yes	No
Docker scout	Yes	Yes	Yes

security vulnerabilities, misconfigurations, malware, and any potential security issues. It scans both OS and non-OS packages of docker images. It collects the vulnerability information from National Vulnerability Database (NVD).

(b) Aqua Trivy: Aqua Trivy is a free-to-use image scanning tool. It scans OS and non-OS packages for detecting vulnerabilities and it also scans dockerfile for checking configurations issues. Aqua trivy vulnerability database collects vulnerability information from NVD, Kube Hunter, Debian, and Red Hat.

(c) jFrog Xray: It is a vulnerability analysis tool for docker images, OS and non-OS packages, and binary files. jFrog Xray vulnerability database collects information from National Vulnerability Database (NVD), Common Vulnerabilities Exposure (CVE) database, and security advisories from software vendors and open-source communities.

(d) Docker Scout: Docker Scout is the docker built-in feature that scans docker images, OCI layout directories, OS, Non-OS packages, and tarball archives to find vulnerabilities. Docker scout database collects vulnerability information from multiple sources such as github advisory database, NVD, bitnami vulnerability database, python packaging advisory database, etc.

(e) Docker Scan: Docker scan is the built-in feature of docker to scan docker images for detecting vulnerabilities. It runs on snyk engine and collects vulnerability information from snyk vulnerability database. Its integration with docker makes it the most popular choice.

4.6 Performance evaluation tools

To select container performance monitoring tools, we followed a set of criteria. We choose widely-used, freely-available tools that have been utilized by practitioners or researchers. Free tools are considered due to the limited capabilities of trial versions of paid products. Container performance monitoring tools such as docker stats, cAdvisor, Prometheus, Grafana, sematext, and locust are used in this study. Prometheus is configured to scrape relevant metrics exposed by docker and the application itself. cAdvisor is deployed as a container runtime monitor, collecting low-level resource utilization data. Grafana is used for visualizations and dashboarding purposes, providing real-time performance insights. The monitoring data provides valuable insights into the performance characteristics and resource demands of the containerized environment throughout our experiments.

(a) Docker Stats: It is a built-in command of docker to compute resource usage of docker containers. When we execute the 'docker stats' command it returns the container ID, container name, CPU utilization, memory utilization, network i/o utilization, and disk i/o usage of all the running containers on docker.

(b) cAdvisor: cAdvisor stands for Container Advisor. It was first introduced by Google and now it is a part of cloud native computing foundation (CNCF). This tool supports real-time monitoring of running containers and provides a web-based interface to view performance metrics with the help of visualizations and graphs. It can work on multiple platforms such as docker and kubernetes.

(c) Prometheus: It was first introduced by Sound Cloud and now it is a part of cloud native computing foundation (CNCF). It is a monitoring and alerting tool used to monitor the health and performance of the containers. It stores the collected data in a time series database which enables users to view and analyze historical data over time. This tool also provides query language known as PromQL that allow users to query and analyze the collected data using various function and operators.

(d) Grafana: It was first introduced by Torkel Ödegaard in 2014 and now it comes under Grafana labs. It supports various data sources such as Prometheus, influxdb, and elastic search to create real-time dashboards and visualizations of data. It provides a web-based interface to create and share visualizations and it supports a wide range of visualization types such as graphs, charts, tables, and maps. Grafana dashboards were designed to display key metrics and performance indicators, allowing us to

monitor the system's behavior and identify potential performance bottlenecks.

(e) Locust: It is an open-source load-testing tool used for testing and optimizing the performance of APIs and web applications. This tool can be integrated with other tools such as Prometheus and Grafana for monitoring and visualizing the results of load tests. The load test can be executed with multiple virtual users, which simulates real-world traffic and stress tests the web applications.

(f) Sematext: It is an open-source log monitoring tool that runs as a container and collects resource usage of all the containers. It automatically identifies containers running on docker and then starts the container monitoring process. It enables us to monitor resource usage metrics, logs, and top containers in a single dashboard.

4.7 Hardware implementation details

The proposed model used in this study is trained and evaluated using google colab. Models are generated using the TensorFlow and Keras frameworks. Then, docker desktop is used for model deployment. The server was running on a machine with 192 GB RAM and × 64-based processor.

In summary, the datasets used to evaluate the prediction performance of the FL model were collected from a publicly available metric repository. All datasets contain static code metrics and process metrics. To overcome the data imbalance problem, random oversampling was performed on the datasets. CNN and ANN were used as the base models to generate the FL model. The implementation was done using the python environment. The prediction performance of generated model was evaluated using AUC and MCC metrics. Additionally, open-source tools such as anchore, aqua trivy, jfrog xray, docker scan, and docker scout were used to analyze the security vulnerabilities in the generated model. To evaluate the performance and resource usage of the generated model, open-source tools such as locust, Prometheus, Grafana, cadvisor, sematext, and docker stats were used.

5 Experimental results

In this section, the experimental results are described. Firstly, the FL model is evaluated on 8 datasets. Then, we analyze the results based on performance metrics. Note that AUC and MCC measures are used to evaluate the prediction performance of the model. After this, we generate the flask api for the model and deploy it using docker. Finally, we analyze the security vulnerabilities and performance of the model with the help of open-source tools.

Table 4 represent the experimental results of the federated learning model in terms of AUC and MCC. According to Table 4, the datasets ant 1.6, xalan 2.5.0, xalan 2.6.0, xerces 1.3.0, and xerces 1.4.4 have the best AUC and MCC values when using the FL model with CNN as the base model. Therefore, out of the 8 selected datasets, 5 of them show the best result when using CNN. Additionally, among all the selected datasets, xerces 1.4.4 shows the best AUC and MCC values for both FL models, whether it is using CNN as the base model or ANN as the base model. Based on Table 4, the answer to RQ1 is presented in Sect. 5.1.

5.1 RQ1: what is the difference between the performances of the FL model having CNN as the base model and the FL model having ANN as the base model?

According to Table 4, we observe that the federated learning model having CNN as the base model outperforms the federated learning model having ANN as the base model as in most of the cases CNN has high AUC and MCC values ANN.

From Table 4, we can also analyze that the FL model having CNN as the base model has high AUC and MCC values for the Xerces1.4.4 dataset. So, the model having CNN as the base model is deployed on docker to analyze the security vulnerabilities and to evaluate the performance in terms of resource usage. Note that we have used Anchore, jFrog Xray, Aqua Trivy, Docker Scan, and Docker Scout to analyze the vulnerabilities. All the tools categorize the security vulnerabilities into four types based on severity levels: Critical, High, Medium, and Low. Table 5 show the vulnerability results of all the chosen tools. Anchore detects 313 vulnerabilities, Aqua Trivy

Table 4 AUC and MCC values for CNN and ANN

Project name	CNN		ANN	
	AUC	MCC	AUC	MCC
Ant 1.6	**0.836**	**0.608**	0.696	0.386
Ant 1.7	0.707	0.385	**0.775**	**0.513**
jEdit 4.1	0.752	0.507	**0.873**	**0.691**
Xalan 2.5.0	**0.738**	**0.481**	0.733	0.466
Xalan 2.6.0	**0.787**	**0.576**	0.761	0.544
Xerces 1.2.0	0.724	0.359	**0.838**	**0.429**
Xerces 1.3.0	**0.873**	**0.663**	0.769	0.495
Xerces 1.4.4	**0.956**	**0.907**	0.883	0.761

The bold values indicate that the highest AUC and MCC values among FL having CNN as the base modeland FL having ANN as the base model for the corresponding dataset

Table 5 Vulnerabilities in docker image

Tools	Critical	High	Medium	Low
Anchore	0	7	194	112
Jfrog xray	0	8	15	10
Aqua trivy	0	5	189	114
Docker scan	0	1	31	22
Docker scout	1	7	113	59

Table 7 Number of users and spawn rate

Number of runs	Number of users	Spawn rate
Run 1	200	50
Run 2	500	50
Run 3	1000	50
Run 4	2000	50
Run 5	2000	100
Run 6	1000	100

detects 308 vulnerabilities, Docker Scout detects 180 vulnerabilities, Docker Scan detects 54 vulnerabilities, and jFrog Xray detects 33 vulnerabilities. Based on Table 5, the answer to RQ2 is presented in Sect. 5.2.

5.2 RQ2: how vulnerable is the FL model?

According to Table 5, docker image of the FL model having CNN as the base model is secure as there exists a smaller number of critical and high-severity vulnerabilities in the case of all the selected tools.

During the scanning of docker images, we identified that there are some vulnerabilities that occur frequently in all the tools. Table 6 represents information related to some of the frequently occurring vulnerabilities in all the tools. These vulnerabilities can lead to a chance of denial-of-service attack, high resource consumption, loss of confidentiality, leakage of sensitive information, etc.

A load test is conducted on the flask API of the machine learning model using the locust tool. 6 runs on the locust tool is performed to check the performance. Table 7 shows the details about number of users and spawn rate taken during various runs. The server is running on a machine with 192 GB RAM and × 64-based processor.

Figure 5 shows the results obtained after performing the load test. During the load test initially, we input the number of users and spawn rate. Spawn rate means the speed at which users are created until the given number of users are created. The metrics which are measured during the load test are response time, number of users, and total requests per second. From Fig. 5, it can be analyzed that as the number of users increases with a constant spawn rate of 50 users the response rate increases but when the spawn rate changes from 50 to 100 users per second then the response rate decreases. In Fig. 5, after run 4 response rate decreases and becomes constant after a certain amount of time. From Fig. 5, we can also analyze that in the case of run 4 and run 5, and run 3 and run 6 number of users are same but there is

Table 6 Identified security vulnerabilities

S.No	CVE-ID	Severity	Description
1	CVE-2023-31484	Medium	Improper certificate validation
2	CVE-2023-29491	Medium	Out of bounds write
3	CVE-2023-46136	Medium	Uncontrolled resource consumption leading to denial of service
4	CVE-2022-47696	Medium	Segmentation fault leading to denial of service or other unspecified impacts
5	CVE-2023-0459	Medium	Allows to leak kernel information
6	CVE-2023-32731	High	Sensitive information disclosure
7	CVE-2023-20588	High	Loss of confidentiality due to division by zero
8	CVE-2023-4206	High	Use after free
9	CVE-2023-42753	High	Potential slab out of bound access due to integer underflow
10	CVE-2023-42756	High	Race condition
11	CVE-2023-28322	Low	Information disclosure
12	CVE-2023-38546	Low	Cookie injection with none file
13	CVE-2023-2602	Low	Memory leak
14	CVE-2021-33294	Low	An infinite loop leading to denial of service
15	CVE-2021-39537	Low	Heap based buffer overflow

a change in spawn rate so when we increase the spawn rate, the response rate decreases during run 5 and run 6.

Note that in this study, open-source tools such as cAdvisor, Prometheus, Grafana, sematext, and docker stats have been used to monitor the performance of docker container. A comprehensive monitoring and visualization infrastructure is established by integrating cAdvisor, Prometheus, and Grafana. Container resource monitoring is conducted using cAdvisor (Container Advisor), an open-source tool designed for collecting and analyzing resource utilization metrics of containerized applications. cAdvisor is integrated into docker desktop to monitor CPU utilization, memory utilization, network traffic, and disk usage of the deployed containers. We use Prometheus to collect and store time-series data from our experimental environment.

Prometheus continuously receives data from cAdvisor and stores it. To visualize the data which is stored by Prometheus, we employed Grafana, an open-source data visualization tool.

Over a monitoring period of 6 h, various metrics including CPU utilization, memory consumption, network traffic, and data usage are collected. Data was collected every 10 s. Table 8 shows the resource usage (CPU utilization, memory utilization, network traffic usage, and disk usage) of FL using five monitoring tools. Among all the selected tools, Grafana shows high resource usage with CPU usage at 85%, memory usage at 453 TB, and network traffic at 1.50 MB/s over a monitoring period of 6 h. cAdvisor, Grafana, Prometheus, and docker stats are specifically used for container monitoring. Sematext provides a wide range of observability tools that covers logging, monitoring, and tracing for both applications and infrastructure. However, docker stats does not continuously monitor the container resource usage; it only provides an immediate snapshot of resource usage by the container. According to Table 8, cAdvisor, Prometheus, and Grafana shows high resource usage as compared to docker stats and sematext.

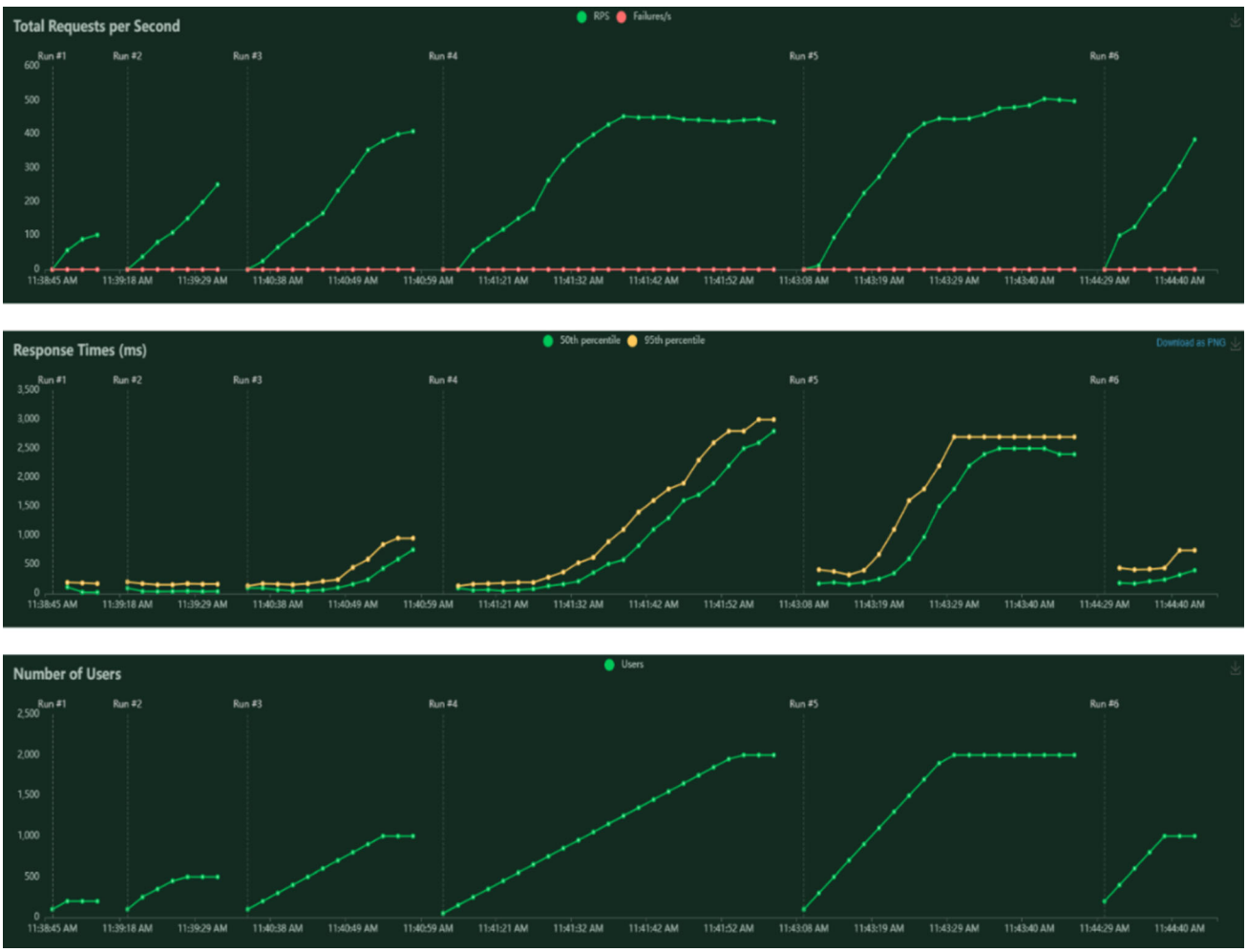


Fig. 5 Load testing using locust

Table 8 Resource usage using monitoring tools

Tools	CPU usage	Memory usage	Network traffic	Disk usage
cAdvisor	80	928.72 MB	317 KB/2389.4 KB	539.81 MB
Docker Stats	146.68	685.4 MB	12.7 KB/53.7 KB	0B/0B
Sematext	137.4	618.6 MB	104.4 KB/2 MB	0B/0B
Prometheus	80	453.20 MB	1.50 KB/s	108 KB
Grafana	85	453 TB	1.50 MB/s	-

Figure 6 illustrates the CPU utilization trend and Fig. 7 illustrates the memory usage trend using Prometheus. Figure 8 illustrates the network traffic trend and Fig. 9 showcases the total bytes used. In Fig. 6, y-axis represents CPU usage percentage, in Fig. 7, y-axis represents memory usage in megabytes, in Fig. 8, y-axis represents bandwidth usage in kilobytes per second, and in Fig. 9, y-axis represents bytes used in kilo bytes. In Figs. 6, 7, 8 and 9, x-axis represents time frame. The results show that after 6 h of run, CPU usage is 80%, memory usage is 453.20 MB, network traffic is 1.50 KB/s, and total bytes used are approximately 108 KB.

Figure 10 illustrates the CPU utilization trend, Fig. 11 illustrates the memory utilization trend, and Fig. 12 showcases the network traffic trend using Grafana dashboards. In Fig. 10, y-axis represents CPU usage percentage, in Fig. 11, y-axis represents memory usage in tera bytes, and in Fig. 12, y-axis represents bandwidth usage in megabytes per second. X-axis represents time frame in Figs. 10, 11 and 12. The result shows that after 6 h of run, CPU usage is approximately 85%, memory usage is 453 TB, and network traffic is 1.50 MB/s.

5.3 RQ3: how does the FL model perform in terms of CPU utilization, memory utilization, network traffic usage, and disk usage?

According to Table 7 and Fig. 5, it can be analyzed that the response rate increases when the spawn rate is 50 users but when the spawn rate increases from 50 to 100 users then the response rate decreases and it can handle multiple users request simultaneously. According to Table 8, Figs. 6, 7, 8, 9, 10, 11, and 12, we can analyze that it performs efficiently as it consumes fewer resources.

The results of this study have implications for practitioners and researchers. The summary of the main findings of this study is as follows: RQ1 examines the prediction performance of FL with CNN as base model and FL with ANN as base model. Based on Table 4, it is observed that FL with CNN as base model achieves higher prediction performance compared to FL with ANN as base model. This is evident from its high AUC and MCC values for most of the selected datasets. RQ2 investigates the security vulnerabilities in the generated model using open-source

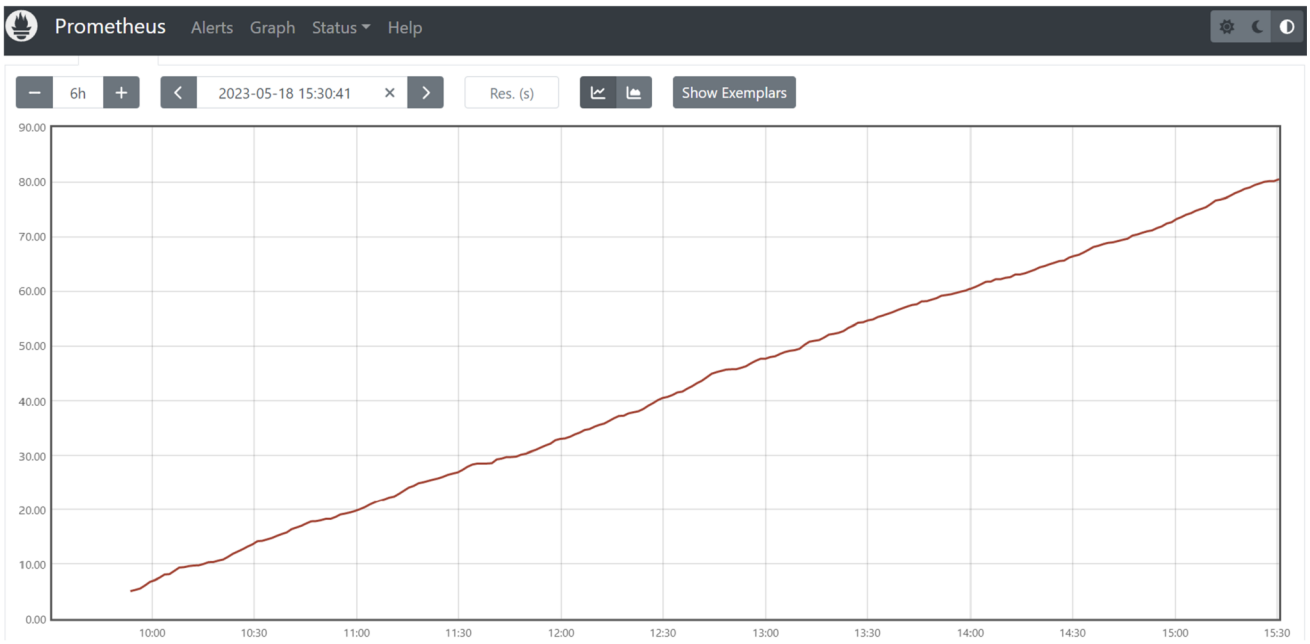


Fig. 6 Prometheus CPU usage



Fig. 7 Prometheus memory usage

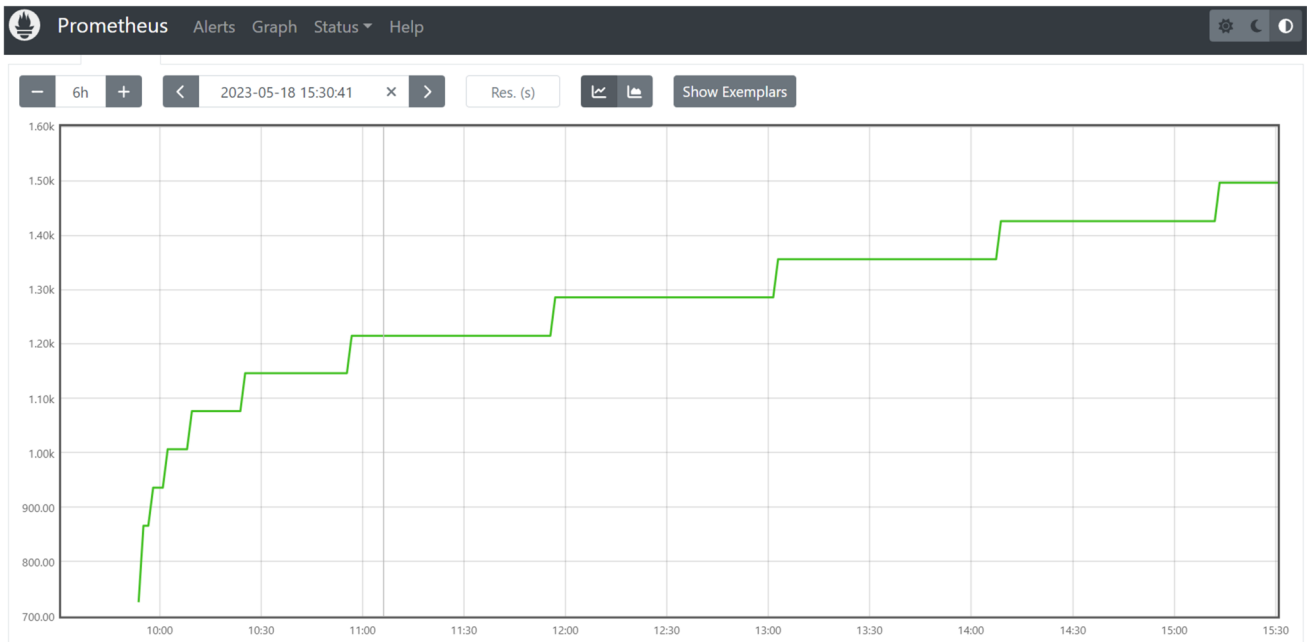


Fig. 8 Prometheus network traffic

tools. Based on Table 5, it is observed that FL model with CNN as base model is deemed secure. This is evident from its lower count of critical and high severity vulnerabilities across most of the selected tools. RQ3 examines the performance and resource usage of generated model in terms of CPU utilization, memory utilization, network traffic usage, and disk usage using open-source tools. Based on Table 7 and Fig. 5, it is observed that as the spawn rate increases from 50 to 100 with a constant number of users, the response rate decreases but the model can handle multiple requests simultaneously. Additionally, based on Table 8 and Figs. 6, 7, 8, 9, 10, 11, 12, it is observed that after 6 h of running, the model consumes fewer resources.

Even though the FL model seems promising, further research is advised to confirm its superiority and investigate possible refinements. This study provides a significant

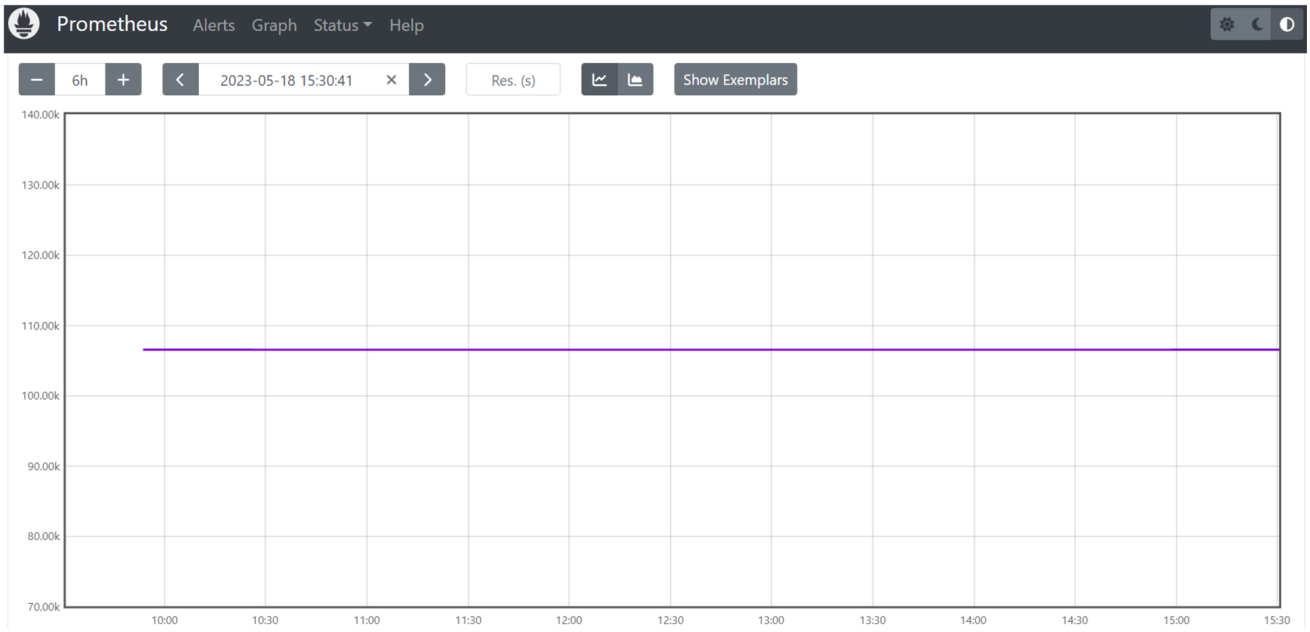


Fig. 9 Prometheus total bytes used

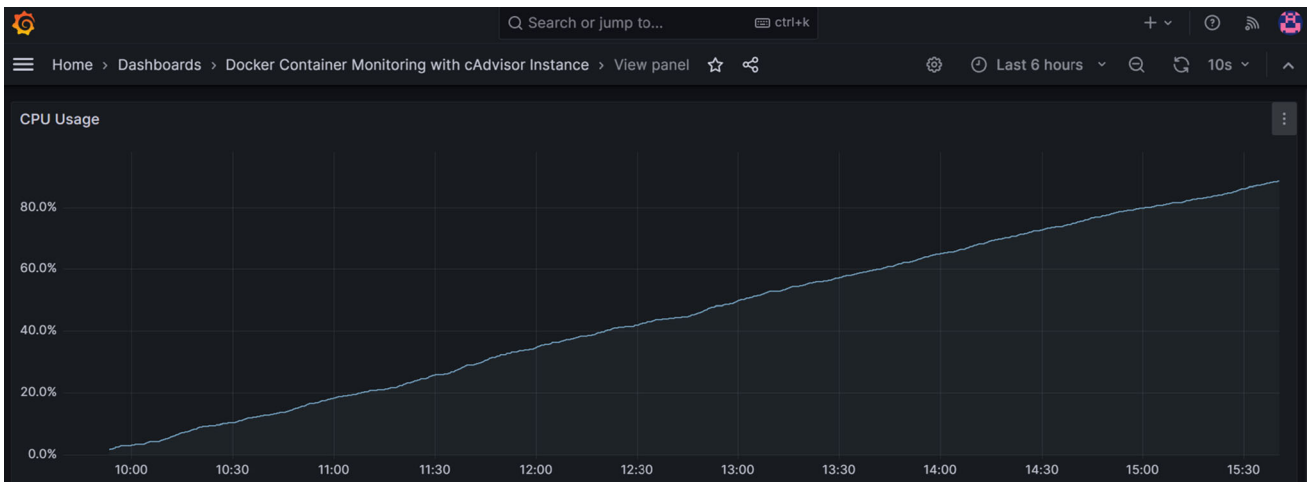


Fig. 10 CPU usage using Grafana dashboard

contribution to the field of SDP and establishes a framework for further research into the efficacy of the FL model and other techniques.

6 Comparison

In this work, 8 distinct datasets are used to find the optimal model among FL with CNN as the base model and FL with ANN as the base model. FL with CNN as the base model performs state of the art compared to FL with ANN as the base model. The performance of these two FL models is compared with ML techniques such as logistic regression (LR), support vector machine (SVM), decision tree (DT), k nearest neighbor (KNN), and naïve bayes (NB). The dataset used in this study is used for the ML techniques. We observe that the proposed model outperforms ML techniques, as shown in Table 9, meanwhile maintaining the privacy and security to the data.

7 Threats to validity

In this section, the threats to our study's validity and the limitation of the study have been discussed. The measures taken to mitigate the threats have also been discussed in this section. It is crucial to evaluate potential threats to validity, including construct, internal, and external factors,

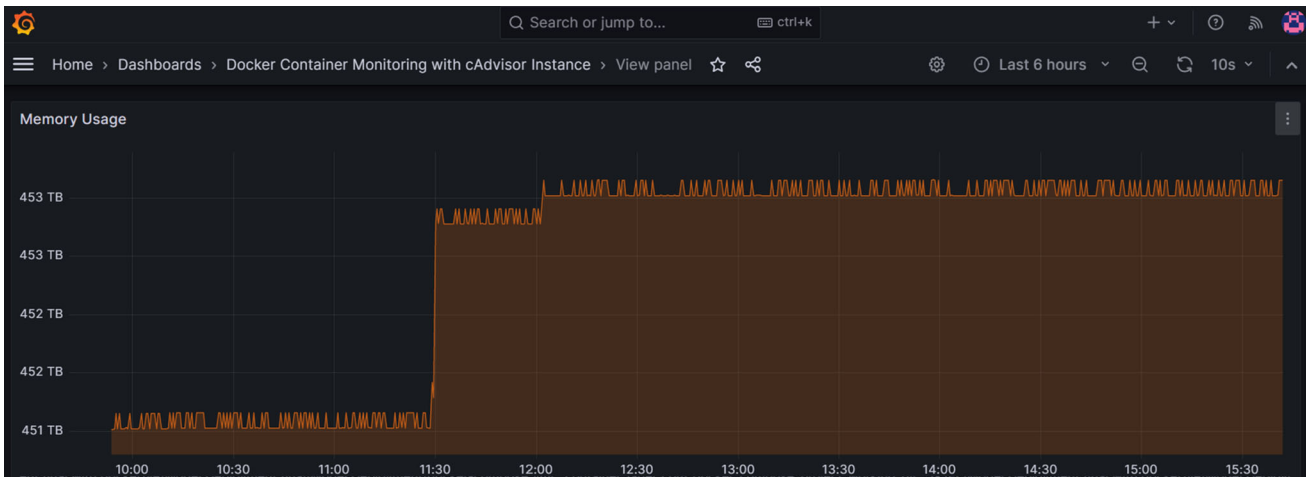


Fig. 11 Memory usage using Grafana dashboard

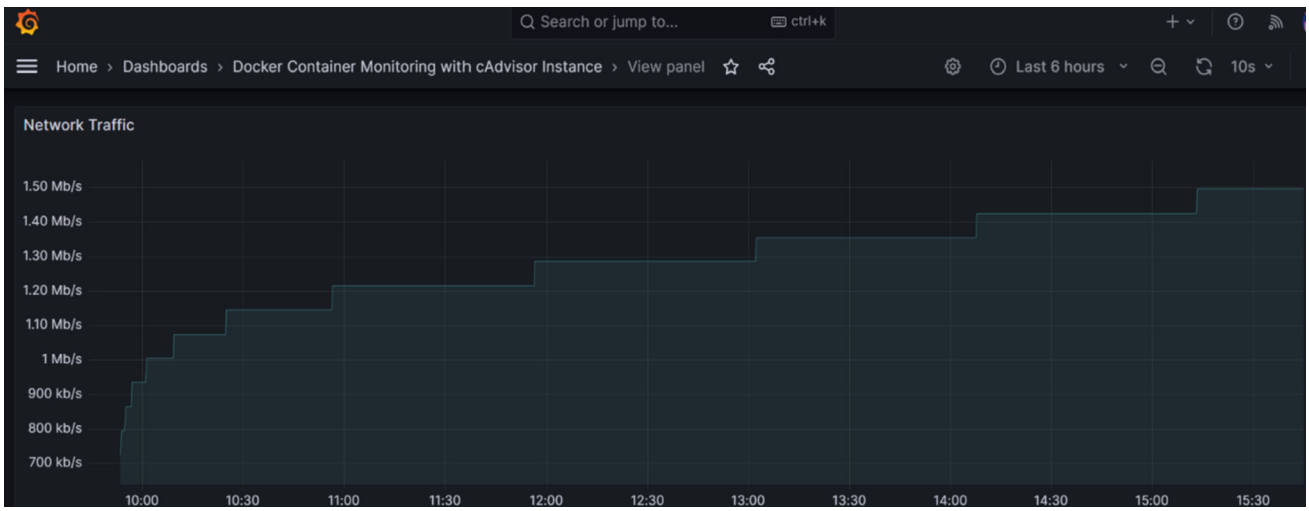


Fig. 12 Network traffic using Grafana dashboard

as well as experiment limitations, especially those related to constraints in the search process and deviations from standard practices.

Internal validity threats are associated with the accuracy of the experiment's outcomes or the procedural aspects of the study. The primary challenge to internal validity revolves around datasets, particularly imbalanced datasets that do not accurately represent the actual distribution of defective and non-defective classes. This threat is mitigated by increasing the classes in original datasets. An additional concern arises from the fact that most of the datasets used in this study have a limited number of defects. The lack of defects poses a challenge in achieving statistically significant results. To resolve this issue, we have employed standard performance measures in this work. However, we acknowledge the potential benefit of using different statistical tests to confirm the statistical significance of our results, and we plan to do so in our future research work.

External validity threats are associated with the generalization of the results of the study. In this study, 8 versions of 4 open-source java projects from publicly available metric repository has been selected for experimentation. We cannot declare that the results of this study can be generalized. It is imperative to acknowledge that future replications of this study are essential to validate and ascertain the broader applicability of our findings.

The limitations of this work are summarized as follows. The datasets used in our work is limited to open-source java projects. Additionally, it is important to note that the results of this study may not be enough for broader generalizations across all software within the industrial domain.

Table 9 Comparison with ML techniques

	Ant 1.6	Ant 1.7	jEdit 4.1	Xalan 2.5.0	Xalan 2.6.0	Xerces 1.2.0	Xerces 1.3.0	Xerces 1.4.4
NB	0.832	**0.889**	**0.926**	0.723	**0.865**	0.737	0.830	0.806
KNN	0.691	0.719	0.746	0.705	0.817	0.700	0.679	0.893
SVM	0.676	0.694	0.740	0.651	0.784	0.540	0.594	0.849
DT	0.813	0.846	0.873	0.730	0.782	0.732	0.763	0.909
LR	0.725	0.726	0.612	0.731	0.833	0.604	0.604	0.853
FL with CNN	**0.836**	0.707	0.752	**0.738**	0.787	0.724	**0.873**	**0.956**
FL with ANN	0.696	0.775	0.873	0.733	0.761	**0.838**	0.769	0.883

8 Conclusion and future work

In this paper, performance evaluation and security analysis of the FL model for software defect prediction is presented by integrating flask, docker desktop, cadvisor, prometheus, sematext, and grafana. FL follows the concept of a distributed learning approach and ensures the privacy of user data. The prediction performance of FL model with CNN as the base model and FL model with ANN as the base model is compared and the results show that CNN outperforms ANN. So, the FL model having CNN as the base model is deployed with the help of docker. The security of the model is analyzed using anchore, aqua trivy, jfrog xray, docker scan, and docker scout. FL model having CNN as base model has a smaller number of critical and high severity vulnerability in case of all the selected tools. It can be concluded that the generated model is more secure as we have used updated packages due to which there are fewer vulnerabilities in the generated model. Locust is used for load testing and the result shows that it can handle multiple users requests simultaneously. A comprehensive monitoring and visualization infrastructure is established by integrating cadvisor, prometheus, and grafana. The resource usage data is collected from the cadvisor dashboard, grafana dashboard, prometheus dashboard, and sematext dashboard and the results show that it consumes fewer resources and performs efficiently.

In the future, we are planning to replicate this work for all classification problems to make this model more generalizable. In other words, we can say that when the docker image of the FL model runs, it should work for all classification problems. We are also planning to use the FL approach for scheduling containers on working nodes. Furthermore, the proposed model can also be used to build defect count model and defect severity model using the process metrics and static code metrics. This work can also be replicated using a more extensive real-life dataset from the industry.

Author contributions RM and MK conceived the presented idea. Implementation was done by AB. The initial draft was written by AB, RM and MK commented on the previous versions of the manuscript. All the authors contributed to the final manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability The datasets used for analysis are available at following links. http://purl.org/MarianJureczko/MetricsRepo

Declarations

Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

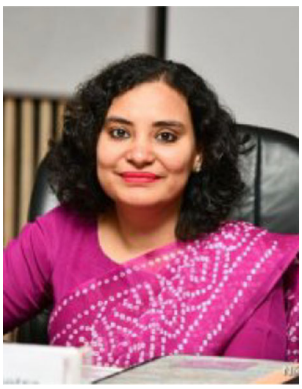
1. Ahmed, M.R., Ali, M.A., Ahmed, N., Zamal, M.F.B., Shamrat, F.M.J.M.: The impact of software fault prediction in real-world application: an automated approach for software engineering. In: Proceedings of 2020 the 6th International Conference on Computing and Data Engineering, in ICCDE 2020, Association for Computing Machinery, New York, pp. 247–251 (2020). https://doi.org/10.1145/3379247.3379278
2. Giray, G., Bennin, K.E., Köksal, O., Babur, O., Tekinerdogan, B.: On the use of deep learning in software defect prediction. J. Syst. Softw.Softw. **195**, 111537 (2023). https://doi.org/10.1016/j.jss.2022.111537
3. Y-cruncher—A Multi-Threaded Pi Program. http://www.numberworld.org/y-cruncher/
4. LINPACK. http://www.netlib.org/linpack/
5. Geekbench 5—Cross-Platform Benchmark. https://www.geekbench.com/
6. Sysbench-GentooWiki. https://wiki.gentoo.org/wiki/Sysbench#Using_the_CPU_workload
7. STREAM Benchmark—AMD. https://developer.amd.com/spack/stream-benchmark/
8. Bonnie++−Wikipedia. https://en.wikipedia.org/wiki/Bonnie%2B%2B
9. Fio—Flexible I/O tester rev. 3.27—fio 3.27-169-g1953e1-dirty documentation. https://fio.readthedocs.io/en/latest/fio_doc.html

10. Netperf—networking benchmark software—LinuxLinks. https://www.linuxlinks.com/netperf/
11. iPerf—The TCP, UDP and SCTP network bandwidth measurement tool. https://iperf.fr/
12. Docker stats | Docker Documentation. https://docs.docker.com/engine/reference/commandline/stats/
13. Cloud Native and Secure DevOps Platform | Sysdig. https://sysdig.com/platform-architecture/
14. Container Monitoring | Datadog. https://www.datadoghq.com/product/container-monitoring/
15. Container Monitoring Tools | Sematext. https://sematext.com/container-monitoring/
16. Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., Gao, Y.: A survey on federated learning. Knowl. Based Syst. **216**, 106775 (2021). https://doi.org/10.1016/j.knosys.2021.106775
17. Li, L., Fan, Y., Tse, M., Lin, K.Y.: A review of applications in federated learning. Comput. Ind. Eng.. Ind. Eng. **149**, 106854 (2020). https://doi.org/10.1016/j.cie.2020.106854
18. Niknam, S., Dhillon, H.S., Reed, J.H.: Federated learning for wireless communications: motivation, opportunities, and challenges. IEEE Commun. Mag.Commun. Mag. **58**(6), 46–51 (2020). https://doi.org/10.1109/MCOM.001.1900461
19. Tran, H.V., Kaddoum, G., Elgala, H., Abou-Rjeily, C., Kaushal, H.: Lightwave power transfer for federated learning-based wireless networks. IEEE Commun. Lett.Commun. Lett. **24**(7), 1472–1476 (2020). https://doi.org/10.1109/LCOMM.2020.2985698
20. Ang, F., Chen, L., Zhao, N., Chen, Y., Wang, W., Yu, F.R.: Robust federated learning with noisy communication. IEEE Trans. Commun.Commun. **68**(6), 3452–3464 (2020). https://doi.org/10.1109/TCOMM.2020.2979149
21. Amiri, M.M., Gündüz, D.: Federated learning over wireless fading channels. IEEE Trans. Wireless Commun.Commun. **19**(5), 3546–3557 (2020). https://doi.org/10.1109/TWC.2020.2974748
22. Li, S., Cheng, Y., Liu, Y., Wang, W., Chen, T.: Abnormal client behavior detection in federated learning. Preprint at https://arxiv.org/abs/1910.09933 (2019)
23. Liu, D., Dligach, D., Miller, T.: Two-stage federated phenotyping and patient representation learning. In: Proceedings of the Conference, Association for Computational Linguistics, pp. 283–291 (2019)
24. Lee, J., Sun, J., Wang, F., Wang, S., Jun, C.H., Jiang, X.: Privacy-preserving patient similarity learning in a federated environment: development and analysis. JMIR Med. Inform. **6**(2), e7744 (2018). https://doi.org/10.2196/medinform.7744
25. Aïvodji, U.M., Gambs, S., Martin, A.: IOTFLA: a secured and privacy-preserving smart home architecture implementing federated learning. In: 2019 IEEE Security and Privacy Workshops (SPW), pp. 175–180 (2019). https://doi.org/10.1109/SPW.2019.00041
26. Feng, J., Rong, C., Sun, F., Guo, D., Li, Y.: PMF: a privacy-preserving human mobility prediction framework via federated learning. Proc ACM Interact Mobile Wearable Ubiquitous Technol **4**(1), 1–21 (2020). https://doi.org/10.1145/3381006
27. Wang, X., Han, Y., Wang, C., Zhao, Q., Chen, X., Chen, M.: In-edge AI: intelligentizing mobile edge computing, caching and communication by federated learning. IEEE Netw.Netw. **33**(5), 156–165 (2019). https://doi.org/10.1109/MNET.2019.1800286
28. Qian, Y., Hu, L., Chen, J., Guan, X., Hassan, M.M., Alelaiwi, A.: Privacy-aware service placement for mobile edge computing via federated learning. Inf. Sci. **505**, 562–570 (2019). https://doi.org/10.1016/j.ins.2019.07.069
29. Komalasari, A., Candra, M.Z.C.: Improving defect prediction using combination of software metrics. In: 2022 International Conference on Data and Software Engineering (ICoDSE), pp. 89–94 (2022). https://doi.org/10.1109/ICoDSE56892.2022.9971813
30. Choudhary, G.R., Kumar, S., Kumar, K., Mishra, A., Catal, C.: Empirical analysis of change metrics for software fault prediction. Comput. Electr. Eng.. Electr. Eng. **67**, 15–24 (2018). https://doi.org/10.1016/j.compeleceng.2018.02.043
31. Rhmann, W., Pandey, B., Ansari, G., Pandey, D.K.: Software fault prediction based on change metrics using hybrid algorithms: an empirical study. J. King Saud Univ. Comput. Info. Sci. **32**(4), 419–424 (2020). https://doi.org/10.1016/j.jksuci.2019.03.006
32. Zhou, Y., Wang, J., Wang, Z.: Bearing faulty prediction method based on federated transfer learning and knowledge distillation. Machines **10**(5), 376 (2022). https://doi.org/10.3390/machines10050376
33. Wang, A., Zhang, Y., Yan, Y.: Heterogeneous defect prediction based on federated transfer learning via knowledge distillation. IEEE Access **9**, 29530–29540 (2021). https://doi.org/10.1109/ACCESS.2021.3058886
34. Wang, A., Zhao, Y., Li, G., Zhang, J., Wu, H., Iwahori, Y.: Heterogeneous defect prediction based on federated reinforcement learning via gradient clustering. IEEE Access **10**, 87832–87843 (2022). https://doi.org/10.1109/ACCESS.2022.3195039
35. Wang, A., Zhao, Y., Yang, L., Wu, H., Iwahori, Y.: Heterogeneous defect prediction algorithm combined with federated sparse compression. IEEE Access **11**, 23739–23753 (2023). https://doi.org/10.1109/ACCESS.2023.3253765
36. Nandi, A., Xhafa, F., Kumar, R.: A Docker-based federated learning framework design and deployment for multi-modal data stream classification. Computing **105**, 1–35 (2023). https://doi.org/10.1007/s00607-023-01179-5
37. He, Y., Meng, G., Chen, K., Hu, X., He, J.: Towards security threats of deep learning systems: a survey. IEEE Trans. Software Eng. **48**(5), 1743–1770 (2020). https://doi.org/10.1109/TSE.2020.3034721
38. Xiao, Q., Li, K., Zhang, D., Xu, W.: Security Risks In Deep Learning Implementations. In: 2018 IEEE Security and privacy workshops (SPW), IEEE, pp. 123–128 (2018). https://doi.org/10.1109/SPW.2018.00027
39. Javed, O., Toor, S.: Understanding the quality of container security vulnerability detection tools. Preprint at https://arxiv.org/abs/2101.03844 (2021)
40. Morabito, R., Kjällman, J., Komu, M.: Hypervisors vs. lightweight virtualization: a performance comparison. In: 2015 IEEE International Conference on Cloud Engineering, pp. 386–393 (2015). https://doi.org/10.1109/IC2E.2015.74
41. Kozhirbayev, Z., Sinnott, R.O.: A performance comparison of container-based technologies for the cloud. Futur. Gener. Comput. Syst.. Gener. Comput. Syst. **68**, 175–182 (2017). https://doi.org/10.1016/j.future.2016.08.025
42. Holopainen, M.: Monitoring Container Environment with Prometheus and Grafana. Metropolia University of Applied Sciences, Helsinki (2021)
43. Leppänen, T.: Data Visualization and Monitoring with Grafana and Prometheus. Turku University of Applied Sciences, Turku (2021)
44. Software Defect Prediction. https://madeyski.e-informatyka.pl/tools/software-defect-prediction/
45. Jureczko, M., Madeyski, L.: Software product metrics used to build defect prediction models. Rep. SPR **2**, 2014 (2011)
46. Madeyski, L., Jureczko, M.: Which process metrics can significantly improve defect prediction models? An empirical study. Softw. Qual. J.. Qual. J. **23**(3), 393–422 (2015). https://doi.org/10.1007/s11219-014-9241-7
47. Dejaeger, K., Verbraken, T., Baesens, B.: Toward comprehensible software fault prediction models using Bayesian network

classifiers. IEEE Trans. Softw. Eng.Softw. Eng. **39**(2), 237–257 (2012). https://doi.org/10.1109/TSE.2012.20
48. Halimu, C., Kasem, A., Newaz, S.H.S.: Empirical comparison of area under ROC curve (AUC) and Mathew Correlation Coefficient (MCC) for evaluating machine learning algorithms on imbalanced datasets for binary classification. In: Proceedings of the 3rd International Conference on Machine Learning and Soft Computing, pp. 1–6 (2019). https://doi.org/10.1145/3310986.3311023
49. Malhotra, R., Bansal, A., Kessentini, M.: Vulnerability analysis of docker hub official images and verified images. In: 2023 IEEE International Conference on Service-Oriented System Engineering (SOSE), pp. 150–155 (2023)

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.



Ruchika Malhotra is Head of Department and Professor in the Department of Software Engineering, Delhi Technological University, Delhi, India. She has served as Associate Dean in Industrial Research and Development, Delhi Technological University. She was awarded with prestigious Raman Fellowship for pursuing Post-doctoral research in Indiana University Purdue University Indianapolis USA. She received her master's and doctorate degree in software engineering from the University School of Information Technology, Guru Gobind Singh Indraprastha University, Delhi, India. She has received IBM Faculty Award 2013. She has been ranked amongst the World's top 2% scientist by Stanford University report, USA, for her research in the field of "Artificial Intelligence & Image Processing" (Ploss Biology, 2020). She is recipient of Commendable Research Award (in 2018-2022) by Delhi Technological University. Her h-index is 36 as reported by Google Scholar. She is author of book titled "Empirical Research in Software Engineering" published by CRC press and co-author of a book on Object Oriented Software Engineering published by PHI Learning. She has published more than 200 research papers in international journals and conferences. Her research interests are in software testing, improving software quality, statistical and adaptive prediction models, software metrics and the definition and validation of software metrics.



Anjali Bansal received the B.Tech. degree in the Department of Computer Science and Engineering from Dr. A. P. J. Abdul Kalam Technical University, Lucknow, India and the M.Tech. degree in the Department of Software Engineering from Delhi Technological University, Delhi, India. She is currently pursuing a Ph.D. in the Department of Software Engineering from Delhi Technological University, Delhi, India. Her Ph.D. project is concerned with the maintenance of software containers, the application of machine learning and deep learning in different areas such as software containers, software quality, and human activity recognition. Her current research interests are AI, software quality, software containers, predictive modelling, and data analytics.



Marouane Kessentini received the Ph.D. degree from the University of Montreal, Canada, in 2012. He is a full professor, chair of the CSE Department, Oakland University, and director of the NSF IUCRC Center on Pervasive AI. He is a recipient of the prestigious 2018 President of Tunisia distinguished research award, the University of Michigan-Dearborn distinguished teaching award, the University of Michigan-Dearborn distinguished digital education award, the University of Michigan-Dearborn/College of Engineering and Computer Science distinguished research award, 4 best paper awards including and the prestigious IEEE 10 Year Most Influential Paper award (2011–2021), and his AI-based software refactoring invention, licensed and deployed by Fortune 500 companies, and selected as one of the Top 8 inventions with the University of Michigan for 2018 (including the three campuses), among more than 500 inventions, by the UM Technology Transfer Office. He received various multi-million grants from both industry and federal agencies and published more than 180 papers in top journals and conferences. He has extensive collaborations with the industry in different areas related to Edge AI, AI/MLOps, AI and cyber-physical systems, intelligent software bots, etc. He is the co-founder of many workshops, general chair of SSBSE16 and ASE22, and PC chair of MODELS19, SANER 2021, GECCO, etc. He served as a keynote speaker with various venues including ICSR, SSBSE, GECCO, WCCI, etc. He graduated more than 18 PhD students and served as associate editor in 7 journals and PC member of more than 200 conferences.

Bull. Mater. Sci. (2024)47:38
https://doi.org/10.1007/s12034-023-03141-9

© Indian Academy of Sciences

Development of KNNLTS–PVDF-based flexible piezoelectric generator for energy-harvesting application

KOMAL VERMA and RICHA SHARMA*
Department of Applied Physics, Delhi Technological University, New Delhi 110042, India
*Author for correspondence (richasharma@dtu.ac.in)

MS received 28 June 2023; accepted 9 October 2023

Abstract. In the present work, a flexible piezoelectric generator (PEG) based on (Li, Ta, Sb) modified (K, Na) NbO_3 (KNNLTS)/poly(vinylidene fluoride) (PVDF) flexible composite film is fabricated. The prepared films are characterized by X-ray diffraction, scanning electron microscopy and Fourier transform infrared spectroscopy. The addition of KNNLTS in the PVDF polymer matrix has resulted in the enhancement in the ferroelectric, piezoelectric, dielectric and piezoelectric output performance. The maximum generated open circuit peak-to-peak voltage from the PVDF/KNNLTS PEG was 36.58 V, which is almost four times higher than the voltage obtained using pure PVDF. The performance of the PEG is also tested under different human body motions, such as fist beating, elbow bending, and quenching and fist opening, which shows the maximum voltage of 5.48, 4.12 and 1.72 V, respectively. The present work demonstrated that high-performance PEG can be developed by modifying KNN ceramics by suitable dopants that can be used in self-powered and wearable devices.

Keywords. Piezoelectric; lead free; composite film; flexible; ferroelectric; energy harvesting.

1. Introduction

Today, the world is facing serious problems related to the energy crisis and environmental pollution that have grabbed considerable attention of the global scientific community to develop a cleaner and sustainable power source [1,2]. In view of this, piezoelectric generators (PEG) that are capable of converting mechanical energy into useful electrical energy are not only very promising solutions but eco-friendly too. Nowadays, due to the crucial role of nanotechnology, energy devices are gradually reducing in size and they also require small operating power. PEG has attracted great attention owing to their ability to produce voltage in the range of nano and micro-order, which is sufficient for charging small batteries, light-emitting devices (LED) and low-power consumable devices [3,4]. Human body motions such as muscle stretching, joint movement, jogging, walking, and heartbeats produce mechanical vibrations which otherwise get wasted and can be used as biomedical energy sources [2]. Therefore, developing sensitive and efficient energy-harvesting devices such as piezoelectric generators which can convert such kinds of biomedical energies into usable electrical signals will be the source of alternative green energy for powering up various portable electronics that are used daily [5–7].

Piezoelectric energy harvester devices employ piezoelectric materials in which changes in the electric dipole moments are due to the mechanical strain and stored in the form of piezoelectric potential [8]. Piezoelectric materials, in general, can be divided into lead-free and lead-containing piezoelectric materials. Among the wide range of available piezoelectric materials, lead-based materials like $Pb[Zr_xTi_{1-x}]O_3$ (PZT) occupy the central position due to their excellent piezoelectric properties. However, it is a known fact that lead-based materials are not sustainable owing to their toxic nature, which destabilizes almost every organ in the human body. In view of this, there is considerable research interest in exploring non-hazardous systems with good piezoelectric properties [9]. Among the lead-free piezoelectric materials, potassium sodium niobate (KNN) ceramics show great potential by their high curie temperature (420°C), electromechanical coupling factor and remnant polarization value (like 33 $\mu C/Cm^2$) and biocompatibility [10–13].

As such, piezoelectric ceramic materials have been extensively studied for constructing piezoelectric generators, as the output performance of these is mainly determined by the piezoelectric properties of the ceramics [14–16]. However, the brittleness of these piezoelectric ceramic materials limits their utility in wearable power generation applications [17]. For the development of piezoelectric generators with high piezoelectricity yet having suitable mechanical flexibility, one approach is to synthesize composites employing piezoelectric ceramic as filler material in the polymer matrix. In this way, the resulting

Supplementary Information: The online version contains supplementary material at https://doi.org/10.1007/s12034-023-03141-9.

Published online: 17 February 2024

composites are expected to have both: excellent piezoelectric properties of the piezoelectric materials and mechanical flexibility of the polymers [18–20].

Because of their lightweight, high flexibility, non-toxicity, biocompatibility, high piezoelectric voltage constant and cost-effectiveness, piezoelectric polymer poly(vinylidene fluoride) PVDF ($d_{31} = \sim 20$ pCN^{-1} and $d_{33} = \sim -20$ to 33 pCN^{-1}) is of tremendous interest in manufacturing of PEG [2,21,22]. PVDF exists in five distinct crystalline phases: α, β, γ, δ, and ε. Among these, α is the non-polar phase, and β and γ are generally the electroactive phases of the PVDF. Thus, the piezoelectric property of the PVDF polymer is mainly determined by the presence of β and γ phases, and β-phase demonstrates the largest dielectric moment among all the crystalline phases [23,24]. Therefore, increasing the β-phase of the PVDF is an efficient way to improve the piezoelectric performance of PVDF-based actuators, sensors and energy harvesters.

Over the years, KNN ceramics in different forms have been incorporated in the PVDF matrix to optimize the PVDF-based piezoelectric energy harvester devices for higher sensitivity and larger piezoelectric output, thus laying a foundation to enhance the technological aspects of PVDF [25,26]. For instance, Bairagi and Ali [27] developed a silane-modified KNN-incorporated PVDF electrospun fibrous web-based nanogenerator. They found a remarkably high output voltage of ~21 V and output current of ~22 μA, which was attributed to de-agglomeration of the nanofillers and better interaction between PVDF chains and KNN nanorods, resulting from the surface modification of the nanofiller material [27]. Furthermore, Abdullah *et al* [28] fabricated a piezoelectric/triboelectric hybrid nanogenerator using composite film made by incorporating the KNN particles and MWCNT in the PVDF matrix and reported very high values of open-circuit voltage and short-circuit current of 54.1 V and ~29.4 μA, respectively [28]. Bairagi and Ali [29] fabricated the PVDF/KNN nanorods-based nanogenerator and studied the effect of KNN concentration in the PVDF matrix on the device performance and obtained a 3.4 V open-circuit voltage for nanogenerator containing 10 wt% KNN nanorods.

To further enhance the piezoelectric output performance of the KNN-based piezoelectric energy harvester, one of the effective approaches is doping the KNN ceramics with suitable dopants to enhance its piezoelectric characteristics, which then lead to superior piezoelectric output performance of the energy harvester device. Saito *et al* [30] in 2004 reported the excellent piezoelectric properties in some KNN-based ceramics. The chemical composition $(K_{0.44}Na_{0.52}Li_{0.04})(Nb_{0.86}Ta_{0.10}Sb_{0.04})O_3$ showed a d_{33} of 416 pC/N, which was significantly higher than the d_{33} of undoped KNN ceramics. Furthermore, Gao *et al* [31] obtained remarkably strong piezoelectricity in KNN-based ceramics doped with Li, Ta and Sb with chemical composition $(K_{0.45}Na_{0.55})_{0.98}Li_{0.02}(Nb_{0.77}Ta_{0.18}Sb_{0.05})O_3$ ceramic (d_{33} = 413 pC/N), which were synthesized by conventional solid-state reaction method [31]. The obtained strong piezoelectric characteristic could be attributed to various factors such as phase coexistence owing to an orthorhombic–tetragonal polymorphic phase transition near room temperature, the higher electronegativity of Sb^{5+} ions in comparison to Nb^{5+} ions and Ta^{5+} ions, high density. Besides, there are many reports available in the literature showing the superior piezoelectric properties of KNN ceramics obtained by doping with Li, Ta and Sb elements in comparison to undoped KNN ceramics [32–34].

However, to the best of our knowledge, studies on the fabrication and characterization of KNNLTS ceramic-based piezoelectric generators have been scarcely reported. For instance, Wang and co-workers [18] developed the composite-type piezoelectric nanogenerator by adding $(K_{0.4425}Na_{0.52}Li_{0.0375})(Nb_{0.86}Ta_{0.06}Sb_{0.08})O_3$ (KNN) particles with Ag particles and CNT in the PDMS matrix and reported a piezoelectric output voltage of ~282 V and short-circuit current of ~32.2 mA. Wang and co-workers [35] studied a fully rollable nanogenerator using the nominal composition of $0.915(Na_{0.5}K_{0.5})(Nb_{0.94}Sb_{0.06})O_3-0.045LiTaO_3-0.04BaZrO_3$ (NKNS-LT-BZ) as a filler material in the PVDF polymer matrix. The fabricated device generated a short-circuit current of 2.6 μA and an open-circuit voltage of 18 V on the application of periodical agitation at a compressive force of 50 N and 1 Hz. In another study, Gupta *et al* [36] developed a piezoelectric energy harvester device using $Na_{0.47}K_{0.47}Li_{0.06}NbO_3$ mixed with polydimethylsiloxane (PDMS) polymer and obtained a large recordable piezoelectric output voltage of 48 V under the vertical compressive force of 2 kgf.

Based on the above literature report, in the present work, we have prepared flexible PVDF/KNNLTS composite film (combining the PVDF and calcined powder of KNNLTS) by the drop cast method. Subsequently, the piezoelectric generator device was constructed by placing the PVDF/KNNLTS composite film between two aluminium tapes. The structural and morphology of the PVDF, KNNLTS ceramics and composite film were done by X-ray diffraction (XRD), scanning electron microscopy (SEM) and Fourier transform infrared spectroscopy (FTIR). Further, the dielectric behaviour, ferroelectric and piezoelectric properties of the PVDF/KNNLTS composite film were investigated. Finally, the piezoelectric output performance of the developed piezoelectric generator was also measured by applying a force on the device surface using a vibrator shaker.

2. Experimental

2.1 *Materials*

Potassium carbonate (K_2CO_3, 99%), sodium carbonate (Na_2CO_3, 99.8%), lithium carbonate (Li_2CO_3, 99%), niobium oxide (Nb_2O_5, (99%), tantalum oxide (Ta_2O_5, 99.95%), and antimony oxide (Sb_2O_5, 99%) purchased from Alfa Aesar were used as raw materials for the synthesis of

$(K_{0.45}Na_{0.55})_{0.98}Li_{0.02}(Nb_{0.77}Ta_{0.18}Sb_{0.05})O_3$ (KNNLTS) ceramic powder. *N*,*N* dimethylformamide (DMF) and the poly(vinylidene fluoride) (PVDF) powder were purchased from Thermo Fischer and Alfa-Aesar, respectively. All of the chemicals were used just as received, with no further treatment.

2.2 *Synthesis of KNNLTS ceramic powder*

To synthesize KNNLTS ceramic powder, a traditional solid-state reaction process was used, in which all of the precursors were first mixed in stoichiometric proportions using propanol as a liquid media by ball milling for 10 h. The obtained homogeneous mixture was dried in a hot air oven at 60°C for 24 h. After drying, the powder was calcined in the furnace at 1050°C for 4 h to obtain KNNLTS ceramic powder. The obtained calcined KNNLTS ceramic powder was then grounded for further use in preparation of composite film.

2.3 *Fabrication of PVDF/KNNLTS flexible composite film*

The procedure of the fabrication of PVDF and PVDF/KNNLTS-based flexible composite films is depicted in figure 1a. The concentration of KNNLTS ceramic powder in the PVDF matrix was fixed at 15 wt%. PVDF powder was dissolved in DMF using a magnetic stirrer for 1 h at 40°C to prepare a uniform transparent solution. Simultaneously, KNNLTS powder was added in DMF and ultrasonicated for 30 min at room temperature. Finally, the two solutions were mixed and stirred using a magnetic stirrer at 50°C for 4 h followed by ultrasonication for 20 min. The prepared homogeneous composite solution was then drop-cast onto a clean glass substrate followed by drying in a hot-air oven at 80°C for 1 hour. Finally, the films were peeled off from the glass substrate to get the PVDF/KNNLTS 15wt% flexible composite films with a thickness of $\sim$40 μm.

2.4 *Fabrication of PVDF/KNNLTS-based piezoelectric generator (PEG)*

To construct the PEG, prepared flexible composite films were cut into small pieces of 2×2 cm^2 and aluminium tape was attached on both sides of the films to make top and bottom electrodes. Then, copper wires were attached to the aluminium tape by using silver paint for the measurement of the generated output voltage and current. Figure 1b shows the schematic diagram of the fabricated PEG device.

2.5 *Characterizations and measurements*

Firstly, the structural analysis of the KNNLTS ceramic powder, pure PVDF film and the PVDF/KNNLTS flexible composite films was done using an X-ray diffractometer (Rigaku, Ultima-1V) with a Cu Kα radiation source (1.54 Å). Fourier transform infrared spectroscopy (FTIR) was used to investigate the *β*-phase formation of PVDF in prepared pure and composite films (Perkin Elmer FTIR spectrum-II). The morphology of synthesized KNNLTS ceramic powder and flexible composite films was studied using scanning electron microscopy (NOVA SEM). The polarization *vs.* electric field (*P vs. E*) hysteresis loop was also traced on Marine India Instrument. Furthermore, the dielectric measurement as a function of frequency was carried out using a Broadband Dielectric/Impedance Analyzer (Model: C-50 Alpha A; Make: Novocontrol Germany, in a frequency range of 20 Hz to 2 MHz). Further, using the electro dynamic shaker (Micron 0020), an external force was applied to the piezoelectric generator, and generated piezoelectric voltage responses were measured using a digital storage oscilloscope (Tektronix, MDO500).

3. Results and discussions

3.1 *Structural analysis*

3.1a *Structural analysis of KNNLTS ceramic powder:* The calcined KNNLTS ceramic powder was examined by XRD and the associated plot is shown in figure 2a. From the plot, it can be seen that the calcined KNNLTS powder exhibits a typical perovskite structure. In addition, no secondary or impurity diffraction peaks are observed in the XRD plot suggesting that the KNNLTS ceramic powder has been successfully synthesized by the solid-state reaction route. The enlarged XRD plot in the 2θ range of 44.5°–47° is presented in figure 2b. From the plot shown in figure 2b, peak splitting can be seen around $\sim$ 45.5°. By assessing the relative intensities of (002) and (200) diffraction peaks, the phase structure of the ceramic can be determined. As reported in the literature, an orthorhombic (O) phase has a (002)/(200) diffraction peaks intensity ratio of 2:1, while the pure tetragonal (T) phase has a (002)/(200) diffraction peaks intensity ratio of 1:2. A ratio approaching 1:1 indicates a coexisted crystal structure with both O and T phases [37,38].

3.1b *Structural analysis of pure PVDF film and PVDF/KNNLTS flexible composite film:* To get the crystallographic information of the prepared PVDF and PVDF/KNNLTS 15wt% composite film, XRD analysis was performed and corresponding plots are shown in figure 3. In PVDF, as mentioned earlier, *β*-phase is the most polar phase and thus responsible for its piezoelectric response, whereas α-phase does not show any piezoelectric response. Figure 3a represents the XRD plot of pure PVDF film. In the XRD plot of pure PVDF film, the diffraction peak at 18.4° can be assigned to the non-polar α-phase of PVDF, whereas the peak at 20.2° corresponds to the polar *β*-phase of the PVDF

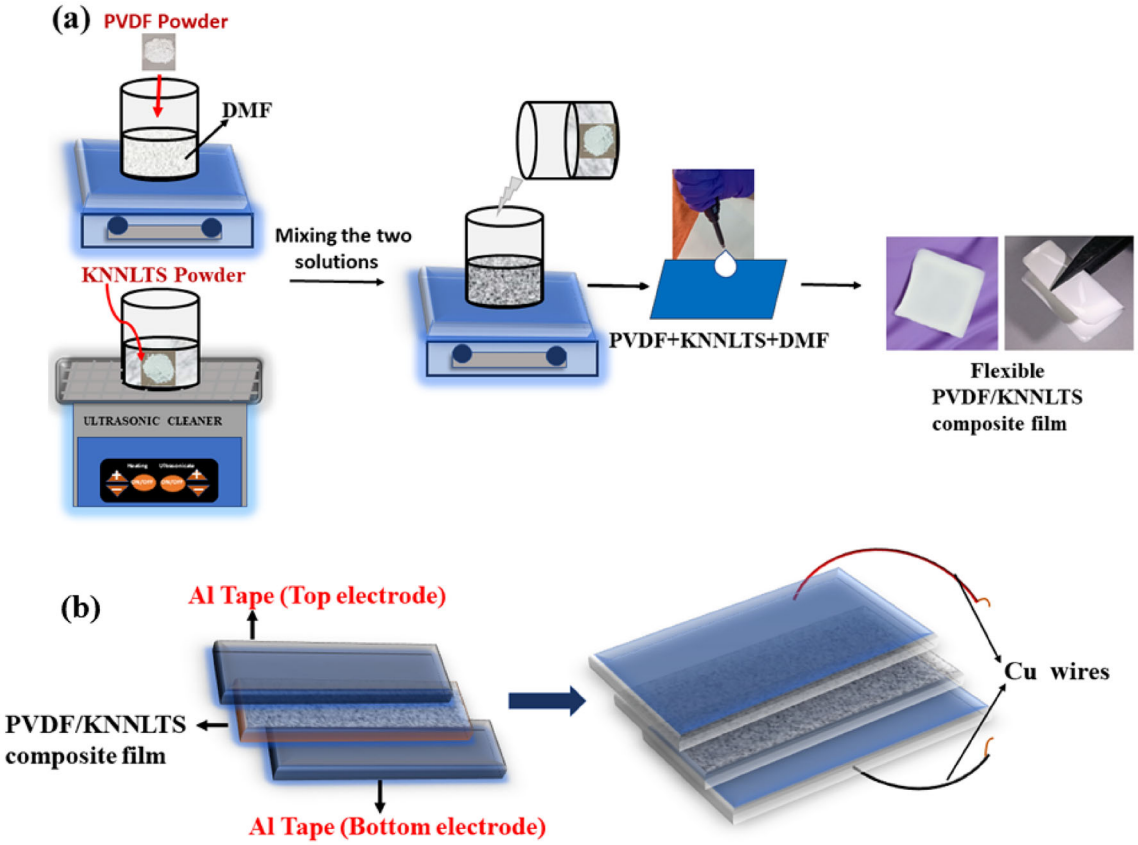


Figure 1. (**a**) Schematic representation of fabrication process of PVDF/KNNLTS flexible composite film. (**b**) Schematic diagram of the fabricated PEG device.

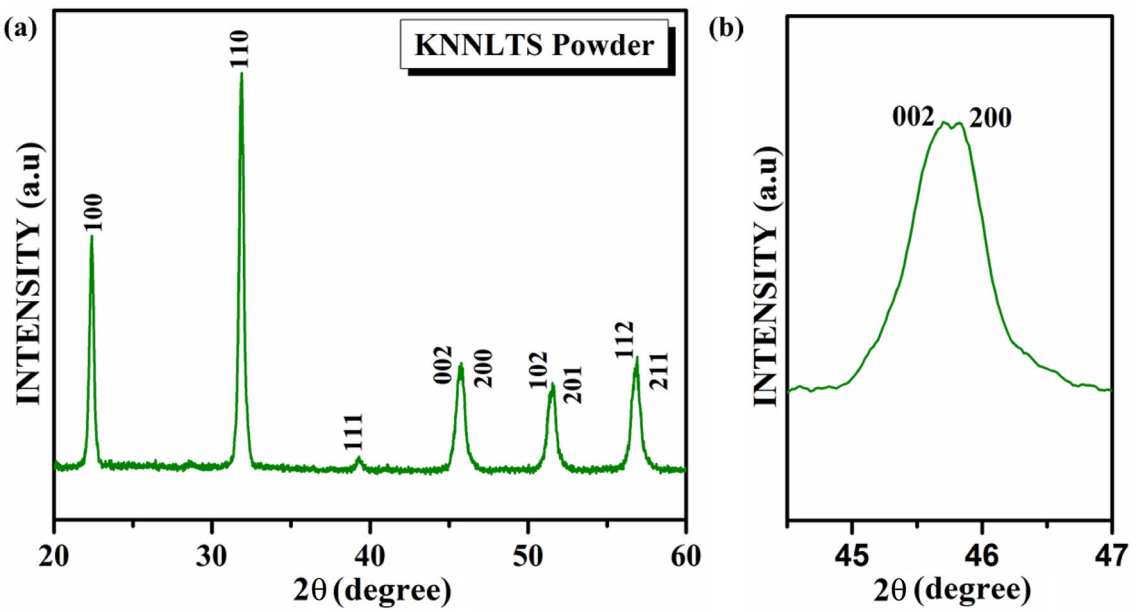


Figure 2. (**a**) XRD pattern of calcined KNNLTS ceramic powder and (**b**) enlarged XRD pattern in the 2θ range of 44.5°–47°.

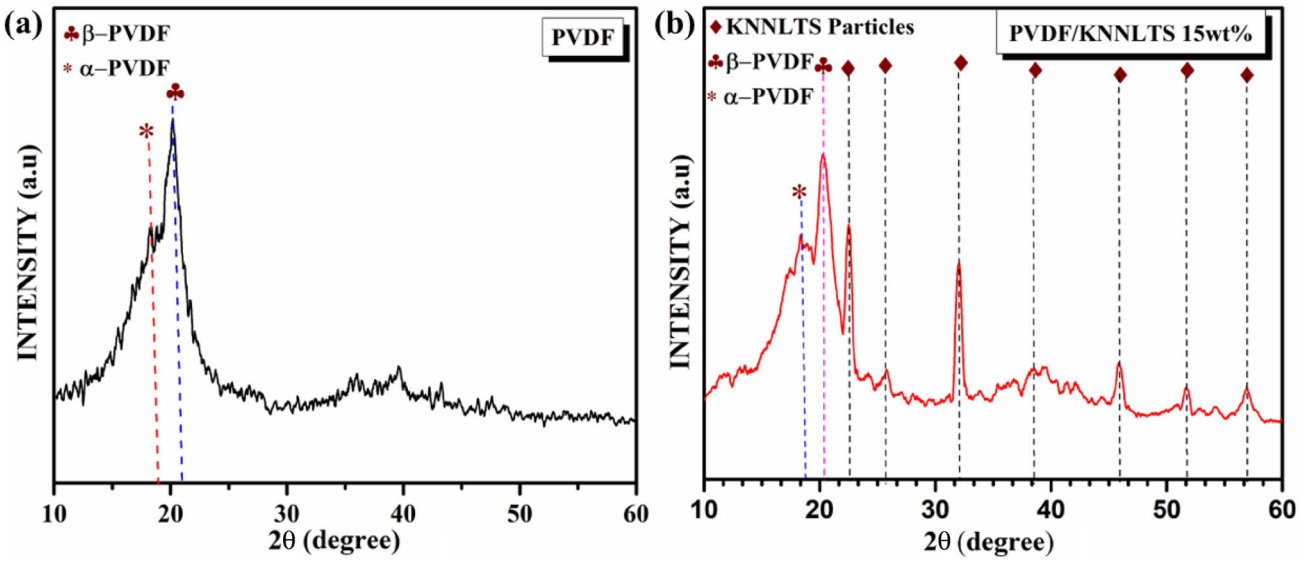


Figure 3. X-ray diffraction pattern of (**a**) PVDF film and (**b**) PVDF/KNNLTS 15wt% flexible composite film.

polymer [39–41]. Additionally, the stronger intensity of the β-phase peak compared to the α-phase suggests that β-phase predominates over α-phase. Furthermore, all the peaks depicted in XRD plot in figure 3b of the PVDF/KNNLTS 15wt% composite film either corresponds to the α and β phases of the PVDF or KNNLTS ceramics powder [31,42].

3.2 *Microstructural analysis of KNNLTS ceramic powder, pure PVDF film and PVDF/KNNLTS composite film*

The morphology of as-synthesized calcined KNNLTS ceramic powder was analysed by SEM and the corresponding image is shown in figure 4. It was observed from the SEM image that KNNLTS ceramic particles exhibited square or rectangular-shaped particles. Further, the surface morphology of the prepared pure PVDF and PVDF/KNNLTS flexible surface film was analysed by using SEM and corresponding images are presented in figure 5a and b. For pure PVDF film, a smooth and dense surface can be seen in the SEM image. From figure 5b, it can be seen that KNNLTS ceramic particles are dispersed uniformly throughout the PVDF, resulting in a uniform film. On the surface of the composite film, KNNLTS particles were not seen thereby confirming the insertion of the ceramic particles in PVDF matrix.

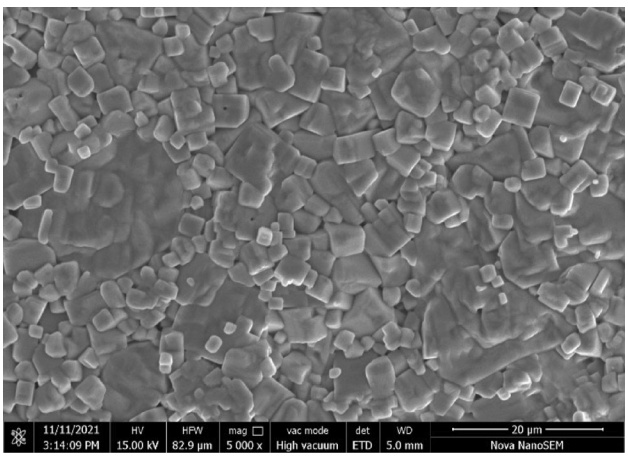


Figure 4. SEM image of KNNLTS calcined ceramic powder.

3.3 *Fourier transform-infrared (FTIR) spectroscopy studies*

The FTIR absorption spectra of pure PVDF film and PVDF/KNNLTS 15wt% composite film were recorded in the range of 600–1500 cm^{-1} at room temperature and are presented in figure 6a and b. The peaks corresponding to the α- and β-phase of the PVDF are marked in figure 6 [39,41,43,44]. The obtained FTIR data was in agreement with the XRD data, depicting the existence of the β-phase of the PVDF in pure PVDF film and PVDF/KNNLTS 15wt% composite film. The Lambert-Beer law can be used to calculate the quantitative value of the β-phase ($F(\beta)$) by using equation (1) [43].

$$F(\beta) = \frac{A_\beta}{\left(\frac{K_\beta}{K_\alpha}\right)A_\alpha + A_\beta} \tag{1}$$

where $K_\beta = 7.7 \times 10^4$ and $K_\alpha = 6.1 \times 10^4$ cm^2 mol^{-1} are the absorption coefficients at 840 and 762 cm^{-1}, respectively [45], A_α and A_β are the absorbance intensities at 762 and 840 cm^{-1}, respectively [39,46]. The values of $F(\beta)$ of PVDF film and PVDF/KNNLTS 15wt% composite film were 43.38 and 52.5%, respectively, indicating the enhancement in the β-phase content due to the addition of KNNLTS ceramic powder in the PVDF matrix.

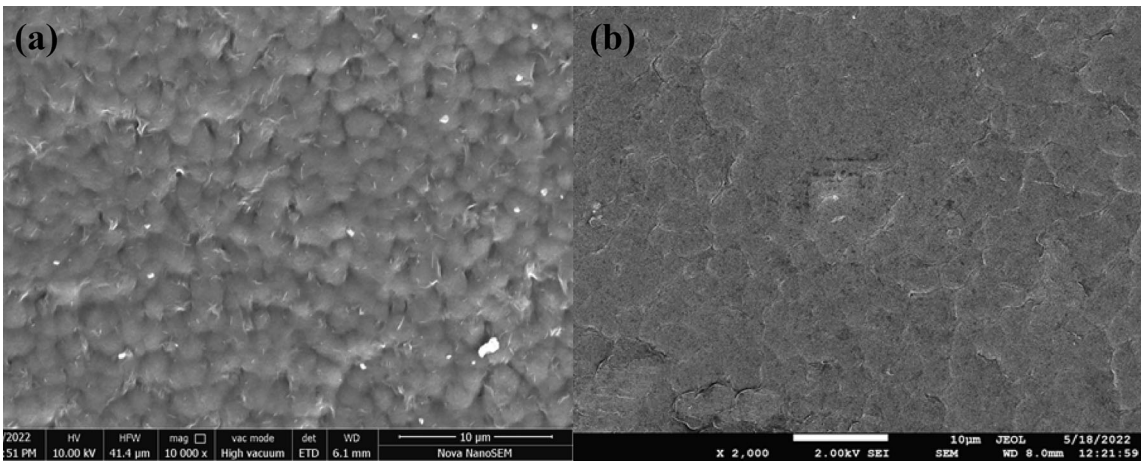


Figure 5. SEM surface micrographs of (**a**) pure PVDF film and (**b**) PVDF/KNNLTS 15wt% composite film.

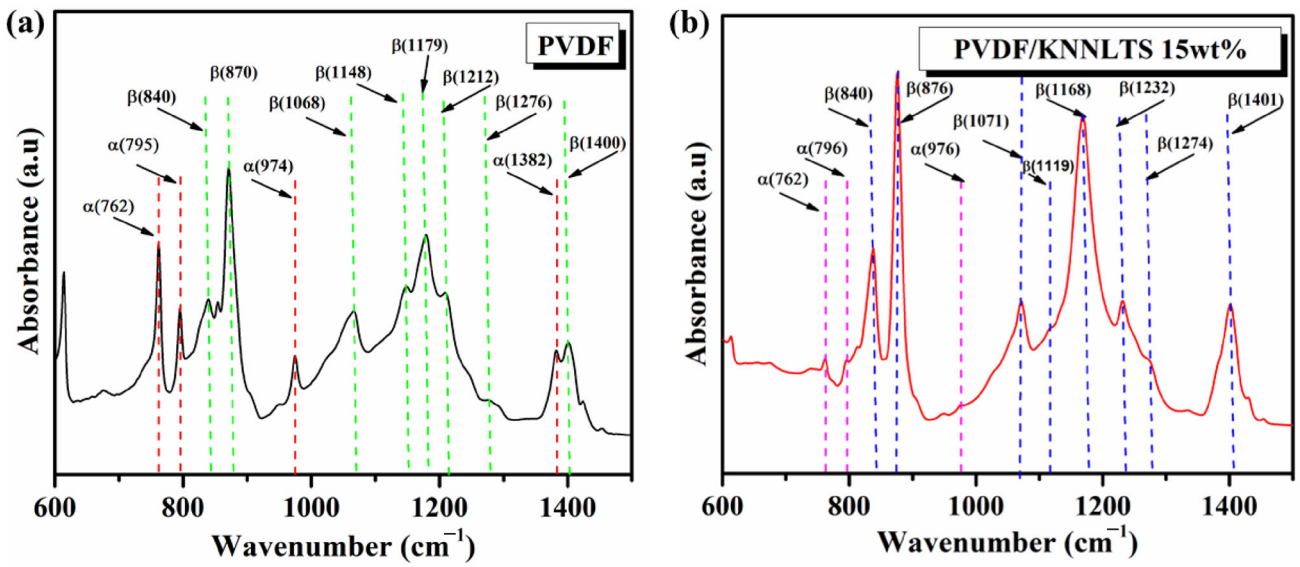


Figure 6. FTIR spectra of (**a**) PVDF film and (**b**) PVDF/KNNLTS 15wt% composite film.

3.4 *Electrical properties*

3.4a *Dielectric properties:* For studying the electrical properties of the prepared PVDF/KNNLTS 15wt% flexible composite film, dielectric constant (ε_r) and dielectric loss (tan δ) were measured. The variation of ε_r and tan δ of prepared composite film as a function of frequency in the range from 20 Hz to 200 kHz is illustrated in figure 7. The dielectric constant of the PVDF/KNNLTS 15wt% flexible composite film decreases with an increase in the frequency which is attributed to a decrease in polarization with an increase in frequency. The polarization is the sum of dipolar, ionic, interfacial and electronic polarizations. When the frequency is low, different polarization responds easily, but with an increase in frequency, different polarization mechanisms cease. This results in the decrease of net polarization of the material, which then leads to a decrease in the dielectric constant [47]. In addition, the dielectric properties are dependent on the content of β-phase and the interfacial polarization taking place due to the addition of KNNLTS ceramic particles in the PVDF matrix. For PVDF/KNNLTS 15wt% composite film, ε_r and tan δ values are 33.56 and 0.0092, respectively, at 1 kHz, which is in good accordance with the dielectric constant values reported for KNN-based ceramic composites in the literature [21,48]. The obtained high value of the dielectric constant will lead to enhanced piezoelectric constant and improved piezoelectric performance of the prepared composite film [49].

3.4b *Ferroelectric property:* The polarization (P) *vs.* the electric field (E) hysteresis loop measurement was done for validating and analysing the ferroelectric properties of the prepared PVDF/KNNLTS 15wt% flexible composite film at room temperature and the corresponding plot is shown in figure 8a. For tracing the loop, a maximum field of

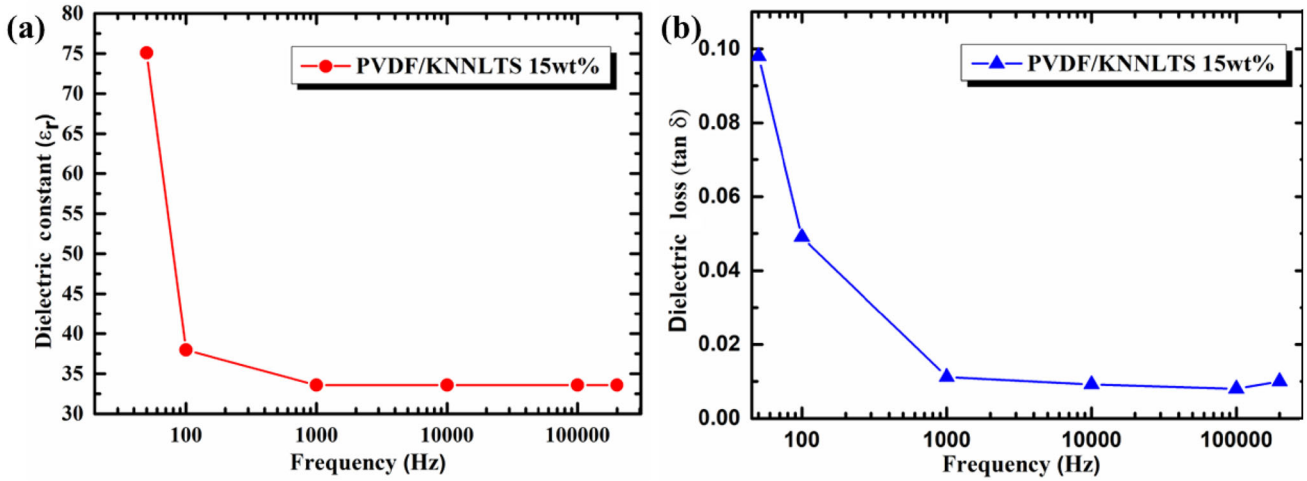


Figure 7. (**a**) Dielectric constant (ε_r) and (**b**) dielectric loss (tan δ) of PVDF/KNNLTS 15wt% composite film as a function of frequency.

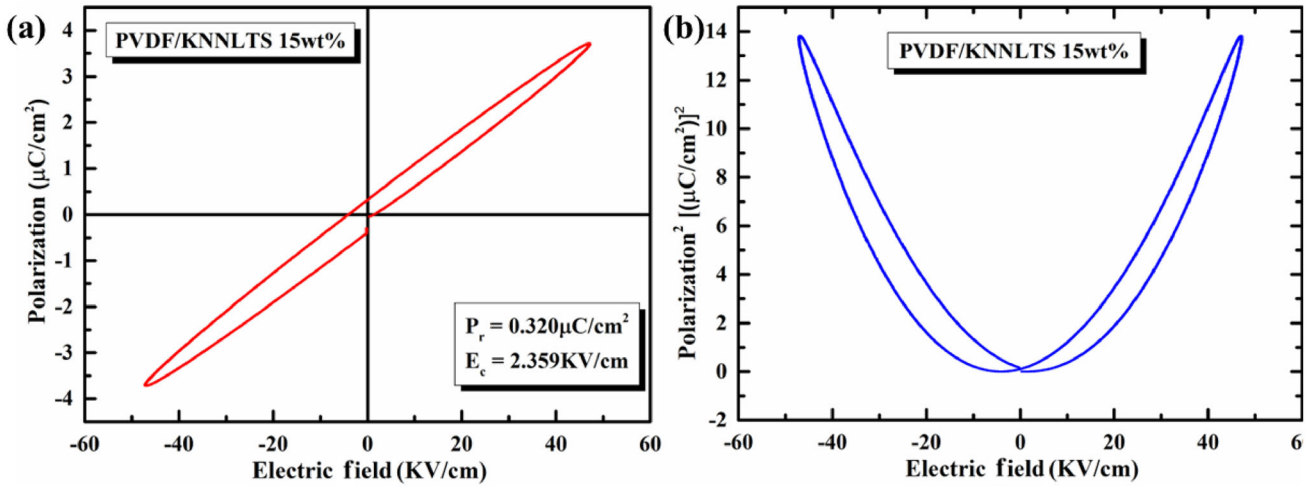


Figure 8. (**a**) Polarization (*P*) *vs.* electric field (*E*) ferroelectric hysteresis loop and (**b**) plot of (polarization)^2 versus the electric field at room temperature for PVDF/KNNLTS 15wt% composite film.

Table 1. Comparison of the output performance of the PEG device developed in the present work with other reported PEG devices.

Piezoelectric filler	Matrix/material	Processes	Output voltage	Mode of pressing	References
KNN nanoblocks	PVDF	Hot-injection	2.1 V	Finger tapping	[21]
KNN nanorods	PVDF	Drop-casting	3.4 V	Finger tapping	[29]
KNN-$BiScO_3$ ceramic particles	Polyimide	Drop-casting	4 V	Finger tapping	[53]
KNN nanofibre	PVDF	Electro-spinning	1.9 V	Thumb tapping	[54]
KNN nanofibre	PVDF	Melt-spinning	17.5 V	Finger tapping	[55]
KNN/ZnO nanorods	PVDF	Electro-spinning	25 V	Finger tapping	[56]
KNN/CNT	PVDF	Electro-spinning	23.24 V	Finger tapping	[57]
KNN-ZS nanofibre	PVDF	Electro-spinning	25 V	Finger tapping	[58]
NKNS-LT-BZ nanoparticles	PVDF	Electro-spinning	18 V	Mechanical stimulator	[35]
$BaTiO_3$ nanoparticles	PVDF	Spin-coating	6.7 V	Finger tapping	[59]
NBT-BT nanoparticles	PVDF	Hot-pressing	9 V	Finger tapping	[60]
KNN ceramic particles	PVDF	Drop-casting	11.2 V	Finger tapping	[52]
KNNLTS ceramic particles	PVDF	Drop-casting	36.58 V	Vibrator shaker	Present work

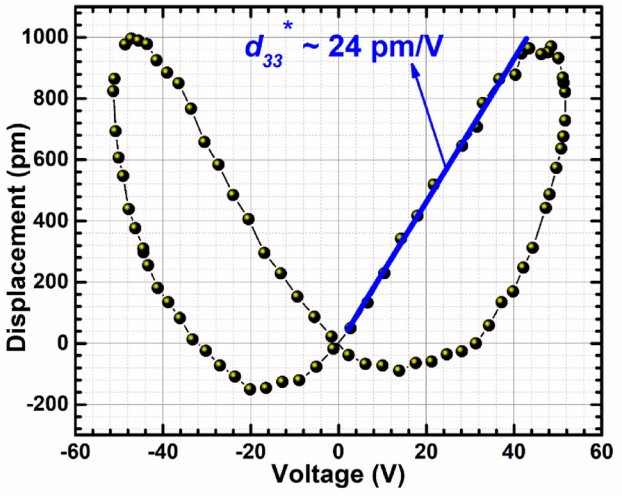


Figure 9. Displacement versus voltage (*D*–*V*) butterfly curve for PVDF/KNNLTS 15wt% composite film.

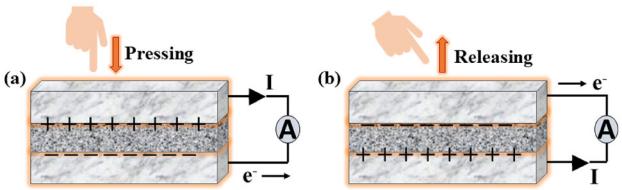


Figure 10. Working mechanism of PEG device.

50 kV cm^{-1} was used. PVDF/KNNLTS 15wt% flexible composite film demonstrated ferroelectricity; however, even after applying a high electric field of 50 kV cm^{-1} to the composite film, polarization cannot reach saturation. From the plot, it can be seen that the polarization increases progressively with an increase in the applied electric field, which is attributed to the better alignment of the electric dipoles at a large field. The measured values of spontaneous polarization (P_s), remnant polarization (P_r), and coercive field (E_c) are 3.707 μC cm^{-2}, 0.320 μC cm^{-2}, and 2.359 kV cm^{-1}, respectively. The obtained values of spontaneous and remnant polarizations in the present work are better than the values reported previously for KNN-based composite energy harvester in the literature [18,29,48].

Polarization switching on the application of an electric field in the piezoelectric material is associated with the piezoelectric strain, which is proportional to (polarization)^2 [16]. Therefore, the plot of (polarization)^2 *vs.* the electric field is equivalent to the strain *vs.* electric field hysteresis plot. In figure 8b, the plot of (polarization)^2 *vs.* electric field for the PVDF/KNLTS 15 wt% composite film is presented. The composite film demonstrated a typical butterfly-shaped loop that is a characteristic of a good piezoelectric system, thereby indicating the potential of PVDF/KNNLTS composite film for piezoelectric energy-harvesting applications.

3.4c *Piezoelectric property:* The presence of piezoelectricity in the prepared PVDF/KNNLTS 15wt% composite film was analysed by measuring the converse piezoelectric coefficient (d_{33}^*). In the present work, the converse piezoelectric coefficient was measured by recording the displacement (*D*)–voltage (*V*) hysteresis loop at room temperature, and the corresponding plot is presented in figure 9. The measured loop for the composite film resembles the shape of a butterfly which is a characteristic of a ferroelectric material. The value of d_{33}^* is obtained from the slope of the butterfly curve [50,51]. The maximum value of d_{33}^* was found to be 24 pm V^{-1}, which is higher than the value of d_{33}^* reported for PVDF/KNN composite film in our previous work [52].

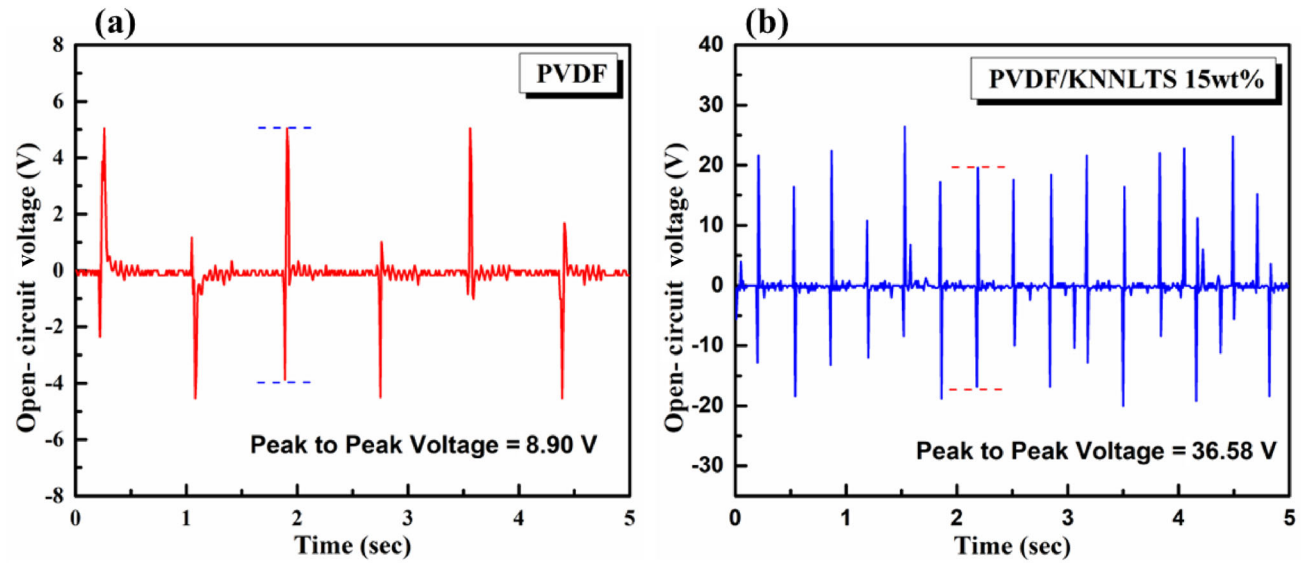


Figure 11. Open-circuit voltage generated using (**a**) PVDF and (**b**) PVDF/KNNLTS 15wt% PEG devices.

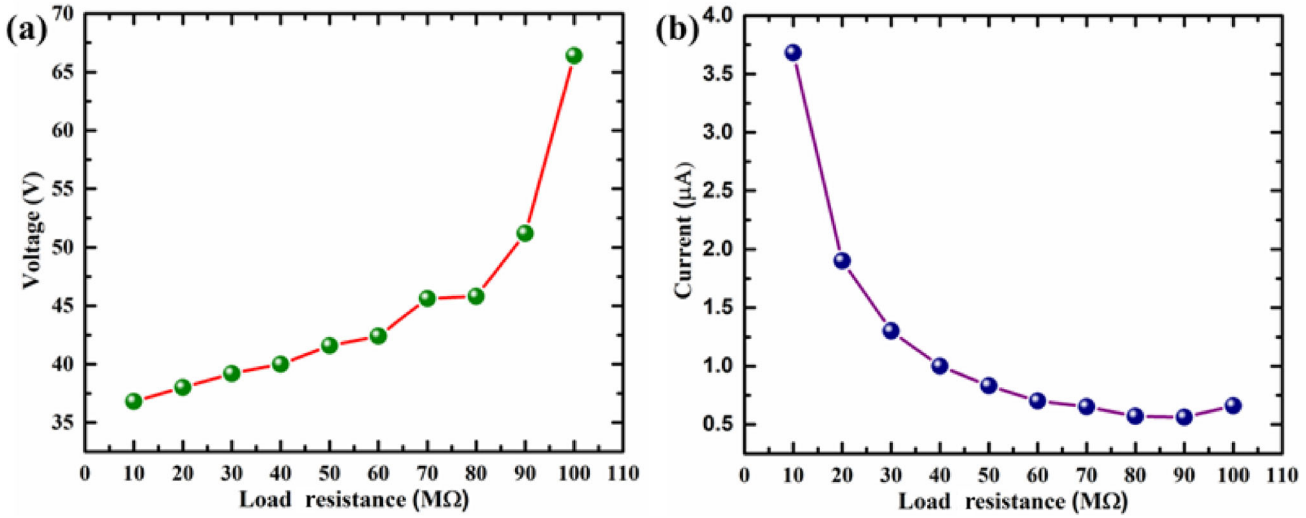


Figure 12. Variation in (**a**) voltage and (**b**) current by varying the load resistance from 10 to 100 MΩ for PVDF/KNNLTS 15wt% PEG device.

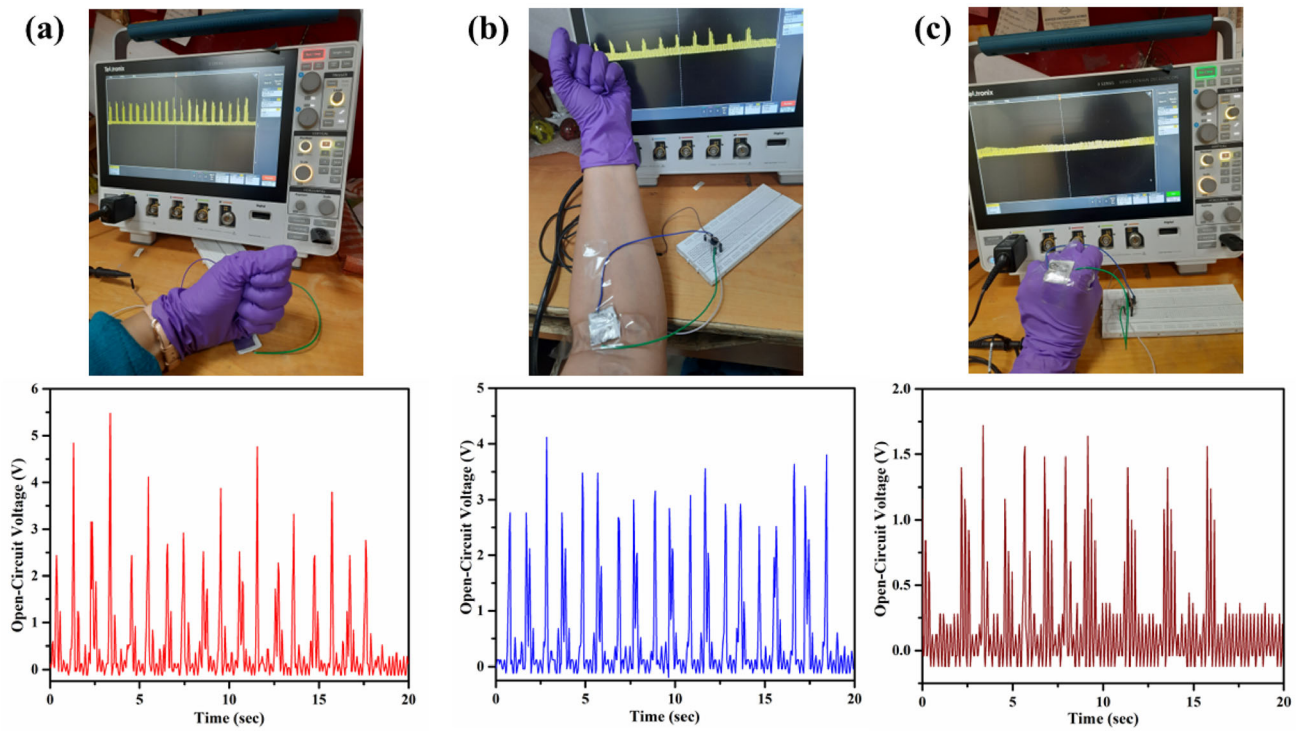


Figure 13. The generated open-circuit voltage when PEG device containing 15wt% KNNLTS ceramic particles in the PVDF matrix for (**a**) fist beating, (**b**) elbow bending, and (**c**) quenching and fist opening.

3.5 *Piezoelectric performance of PEG device*

In figure 10, the working mechanism of the constructed PEG device is shown. Electric dipoles orient themselves in a single direction when a perpendicular force is applied to the device, creating potential differences between the two electrodes. As a result, as shown in figure 10a, current flows through an outer circuit from top to bottom electrodes. Further, when the applied force on the device is released, the electric dipoles cause the current to flow from the bottom to the top electrodes, as presented in figure 10b. The open-circuit output voltage of the constructed PVDF/KNNLTS PEG device with dimensions of the order of 2 × 2 cm by applying a force (∼1.5 N) at a tapping frequency of 3 Hz on its surface was measured using a vibrator shaker device. When pressure is applied to the device's surface, a potential difference develops across the generator, which is then measured by the digital storage

oscilloscope. Figure 11 illustrates the plot of the open-circuit voltage generated by the constructed PEG using pure PVDF film and PVDF/KNNLTS 15wt% composite film. The maximum peak-to-peak open-circuit voltage (V_{P-P}) for the pure PVDF and PVDF/KNNLTS 15wt% based PEG was 8.90 and 36.58 V, respectively, showing that the addition of KNNLTS in the PVDF matrix has led to fourfold enhancement in the peak-to-peak open-circuit voltage of the PEG device. Moreover, the obtained open-circuit voltage of PVDF/KNNLTS 15wt% PEG device was considerably higher than PVDF/KNN-based PEG device as reported in our previous work [52], which is mainly attributed to the superior piezoelectric and polarization characteristics of KNN ceramics, resulting from the doping with Li, Sb and Ta as reported in the literature [31]. Table 1 presents the comparison of the parameters between the PEG device developed in the present work and the other reported PEG devices.

Furthermore, the output performance of the fabricated PEG device was also analysed by applying different load resistances in the range of 10 to 100 MΩ, and corresponding plots are shown in figure 12. From figure 12a and b, it can be seen that as load resistance increased the voltage drop also increased, while the current was decreased with an increase in load resistance. The maximum value of output voltage obtained was around 66.4 V for PVDF/KNNLTS 15wt% PEG device at a load resistance of 100 MΩ.

To demonstrate the ability of the fabricated PEG device in real-time applications, PEG device was placed in different positions on the hand and leg to harvest mechanical energy. Figure 13 shows the PVDF/KNNLTS 15wt% PEG device employed to extract human body motions, such as fist beating (supplementary Video 1), elbow bending (supplementary Video 2), and quenching and fist opening (supplementary Video 3). The maximum open-circuit voltages obtained by fist beating, elbow bending, and quenching and fist opening were 5.48, 4.12 and 1.72 V, respectively. These results prove that the device fabricated in the present work with PVDF/KNNLTS 15wt% composite film has the potential to utilize the various mechanical energy from our daily life.

4. Conclusion

In conclusion, in the present research work, we fabricated eco-friendly PVDF/KNNLTS flexible composite films by the drop cast method. In comparison to pure PVDF film, PVDF/KNNLTS 15wt% composite film showed the presence of a higher fraction of β-phase in the film, which is crucial in improving the piezoelectric output performance of the fabricated PEG device. Furthermore, we were able to achieve superior output voltage signals for the PEG device containing 15wt% KNNLTS ceramic powder as fillers in the PVDF matrix. The maximum open-circuit peak-to-peak voltages produced by the PEG device fabricated with pure PVDF film and with 15wt% KNNLTS in the PVDF matrix were 8.90 and 36.58 V, respectively. The obtained piezoelectric output voltage in the present work is significantly higher than the voltage reported for many PVDF/KNN-based piezoelectric energy harvester devices. Owing to the good mechanical flexibility, biological compatibility, straightforward synthesis procedure and enhanced piezoelectric performance, the fabricated PEG device in the present work demonstrates a lot of potential for self-powered and wearable devices.

Acknowledgements

We wish to thank the DTU Project Research Grant (F.NO.DTU/IRD/619/2019/2113) for providing the research grant to perform the research work. Komal Verma expresses her gratitude to Delhi Technological University (DTU), India, for providing the DTU fellowship (2K19/PHDAP/06), and is also thankful to Dr Sahil Goel for helping in doing data collection and characterization.

References

[1] Chen W and Lei Y 2018 *Renew. Energy* **123** 1
[2] Badatya S, Bharti D K, Srivastava A K and Gupta M K 2021 *J. Alloys Compd.* **863** 158406
[3] Kaur A, Gupta A, Ying C, Rahmani M and Sapra G 2023 *Microelectron. Eng.* **275** 1119992
[4] Kuang Y and Zhu M 2019 *Compos. Part B Eng.* **158** 189
[5] He J, Zhang P, Jing P, Deng L and Zhang H 2023 *Microelectron. Eng.* **275** 111994
[6] Seung W, Gupta M K, Lee K Y, Shin K S, Lee J H, Kim T Y *et al* 2015 *ACS Nano* **9** 3501
[7] Yaqoob U, Habibur R M, Sheeraz M and Kim H C 2019 *Compos. Part B Eng.* **159** 259
[8] Cheng X, Liu Z, Jing Q, Mao P, Guo K, Lu J *et al* 2023 *J. Colloid Interface Sci.* **629** 11
[9] Jiang F, Zhhao W-S and Zhang J 2020 *Microelectron. Eng.* **225** 111279
[10] Wu J, Xiao D and Zhu J 2015 *Chem. Rev.* **115** 2559
[11] Verma K, Goel S and Sharma R 2022 *J. Mater. Sci.: Mater. Electron.* **33** 26067
[12] Zhang Y, Kim H, Wang Q, Jo W, Kingon A I, Kim S-H *et al* 2020 *Nanoscale Adv.* **2** 3131
[13] Egerton L and Dillon D M 2018 *J. Am. Ceram. Soc.* **42** 438
[14] Zheng M, Hou Y, Chao L and Zhu M 2018 *J. Mater. Sci.: Mater. Electron.* **29** 9582
[15] Zhang Y, Liu X, Wang G, Li Y, Zhang S, Wang D *et al* 2020 *J. Alloys Compd.* **825** 154020
[16] Batra K, Sinha N and Kumar B 2020 *J. Alloys Compd.* **818** 152874
[17] Zhou Z, Du X, Zhang Z, Luo J, Niu S, Shen D *et al* 2021 *Nano Energy* **82** 105709
[18] Huan Y, Wei T, Wang Z, Shen H, Lin X, Huang S *et al* 2020 *J. Materiomics* **6** 355

[19] Bairagi S and Ali S W 2019 *Eur. Polym. J.* **116** 554
[20] Wua Y, Maa F, Qua J, Luo Y, Lv C, Guo Q *et al Appl. Surface Sci.* **469** 283
[21] Nair K S, Varghese H, Chandran A, Hareesh U N S, Chouprik A, Spiridonov M *et al* 2022 *Mater. Today Commun.* **31** 103291
[22] Bai Z, Yao Z, Wu G, Liu K, Ye D, Tao Y *et al* 2023 *Sens. Actuators: A Phys.* **357** 114407
[23] Wang X M, Sun F Z, Yin G C, Wang Y T, Liu B and Dong M D 2018 *Sensors* **18** 16
[24] Lu L, Ding W, Liu J and Yang B 2020 *Nano Energy* **78** 105251
[25] Jin L, Ma S, Deng W, Yan C, Yang T, Chu X *et al* 2018 *Nano Energy* **50** 632
[26] Ryu C, Hajra S, Sahu M, Jung S I, Jang I R and Kim H J 2022 *Mater. Lett.* **309** 131308
[27] Bairagi S and Ali S W 2020 *Energy* **198** 117385
[28] Abdullah A M, Sadaf M U K, Tasnim F, Vasquez H, Lozano K and Uddin M K 2021 *Nano Energy* **86** 106133
[29] Bairagi S and Ali S W 2020 *Org. Electron.* **78** 105547
[30] Saito Y, Takao H, Tani T, Nonoyama T, Takatori K, Homma T *et al* 2004 *Nature* **423** 84
[31] Gao Y, Zhang J, Qing Y, Tan Y, Zhang Z and Hao X 2011 *J. Am. Ceram. Soc.* **94** 2968
[32] Lv Y G, Wang C L, Zhang J L, Wu L, Zhao M L and Xu J P 2009 *Mater. Res. Bull.* **44** 284
[33] Ming B-Q, Wang J-F, Qi P and Zang G-Z 2007 *J. Appl. Phys.* **101** 054103
[34] Wang K and Li J-F 2010 *J. Am. Ceram. Soc.* **93** 1101
[35] Zhang C, Fan Y, Li H, Li Y, Zhang L, Cao S *et al ACS Nano* **12** 4803
[36] Gupta M K, Kim S-W and Kumar B 2016 *ACS Appl. Mater. Interfaces* **8** 1766
[37] Dwivedi S, Pareek T and Kumar S 2018 *RSC Adv.* **8** 24286
[38] Zhang M H, Wang K, Zhou J S, Zhou J S, Chu J C, Lv X *et al* 2017 *Acta Mater.* **122** 344
[39] Jaleh B and Jabbari A 2014 *Appl. Surf. Sci.* **20** 339
[40] Devi P and Ramachandran K 2011 *J. Exp. Nanosci.* **6** 281
[41] Hwang H J, Yang J H, Kang S C, Cho C, Kang C G, Lee Y G *et al* 2013 *Microelectron. Eng.* **109** 87
[42] Li Y, Su X, Liang K, Luo C, Li P, Hu J *et al* 2021 *Microelectron. Eng.* **244–246** 111557
[43] Bhatt A, Singh V, Bamola P, Aswal D, Rawat S, Rana S *et al* 2023 *J. Alloys Compd.* **960** 170664
[44] Soin N, Boyer D, Prashanth K, Sharma S, Narasimulu A A, Luo J *et al* 2015 *Chem. Commun.* **51** 8257
[45] Sabry R S and Hussein A D 2019 *Polym. Test.* **79** 106001
[46] Lee J S, Shin K-Y, Cheong O J, Kim J H and Jang 2015 *J. Sci. Rep.* **5** 7887
[47] Chandrasekhar M and Kumar P 2015 *Ceram. Int.* **41** 5574
[48] Mirzazadeh Z, Sherafat Z and Bagherzadeh E 2021 *Ceram. Int.* **47** 6211
[49] Jella V, Ippili S, Eom J-H, Choi J and Yoon S-G 2018 *Nano Energy* **53** 46
[50] Sinha N, Goel S, Joseph A J, Yadav H, Batra K, Gupta M K *et al* 2018 *Ceram. Int.* **44** 8582
[51] Goel S, Sinha N, Yadav H, Joseph A J and Kumar B 2017 *Phys. E Low-Dimens. Syst. Nanostruct.* **91** 72
[52] Verma K and Sharma R 2023 *Chin. J. Phys.* **84** 198
[53] Hyeon D Y, Lee G J, Lee S H, Park J J, Kim S, Lee M *et al* 2022 *Compos. Part B* **234** 109671
[54] Teka A, Bairagi S, Shahadat Md, Joshi M, Ahammad S Z and Ali S W 2018 Polym. *Adv. Technol.* **29** 1
[55] Bairagi S and Ali S W 2019 *Energy Technol.* **7** 1900538
[56] Bairagi S and Ali S W 2020 *Int. J. Energy Res.* 1
[57] Zhang Z, Chen Y and Guo J 2019 *Phys. E: Low-Dimens. Syst. Nanostruct.* **105** 212
[58] Goel S, Sinha N, Yadav H, Joseph A J and Kumar B 2017 *Physica E* **91** 72
[59] Hu P, Yan L, Zhao C, Zhang Y and Niu J 2018 *Compos. Sci. Technol.* **168** 327
[60] Petrovic M V, Cordero F, Mercadelli E, Brunengo E, Ilic N, Galassi C *et al* 2021 *J. Alloys Compd.* **884** 161071

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.

Arabian Journal for Science and Engineering
https://doi.org/10.1007/s13369-024-08705-3

RESEARCH ARTICLE-PHYSICS

Device Optimization of T-shaped Gate and Polarized Doped Buffer-Engineered InAlN/GaN HEMT for Improved RF/Microwave Performance

Megha Sharma[1] · **Rishu Chaujar**[1]

Received: 4 June 2022 / Accepted: 31 December 2023
© King Fahd University of Petroleum & Minerals 2024

Abstract
III-nitride high-electron-mobility transistors (HEMTs) are the optimal choices for high-RF and power applications due to their exceptional performance characteristics. Nevertheless, HEMTs face formidable challenges related to large gate and buffer leakage currents, which inevitably impose significant limitations on device performance and reliability. In response to these challenges, this work studies the design consideration of a T-shaped gate InAlN/AlN with polarized doped buffer HEMT. The simulation results show that the parasitic capacitances are deeply affected by the geometry of the T-gate. The less influence of gate-to-source capacitance (C_{gs}) by the drain voltage possesses constant depletion charges at the source side, leading to a persistent value of C_{gs} with drain bias. Moreover, a polarization-doped buffer layer has been incorporated, effectively enhancing the confinement of the two-dimensional electron gas by causing an upward convex bend in the conduction band. This strategic design element significantly contributes to the reduction in leakage current (10^{-8}), positive shift of V_{th} (0.9 V), and improves the switching ratio (10^8). The effect of gate head length (H_{length}), gate stem height (S_{height}), and gate foot length (F_{length}) on various RF parameters has been investigated. The simulated results confirm that the proper choice of H_{length} (280 nm), S_{height} (100 nm), and F_{length} (10 nm) significantly reduced the parasitic capacitance (C_{gs} = 350 fF/mm and C_{gd} = 140 fF/mm) and enhanced the F_{max} (840 GHz), GBP (636 GHz), F_t (583 GHz) and also improved the power gains. The extracted results of the proposed design exhibit its efficacy for high-performance RF/microwave applications.

Keywords Cutoff frequency · Conduction band · Capacitance · Heterostructure · Polarization · T-shaped gate

1 Introduction

Past two decades, silicon technology has dominated the semiconductor industry. In the current era, silicon-based devices, including FinFETs [1–3], recessed channels [4–6], MOSFETs cylindrical gates [7, 8], and nanowires [9, 10], are steadily approaching their theoretical performance limits and are unable to fulfill the demand of higher-power-RF applications. Gallium nitride (GaN) has favorable characteristics that position it as a promising contender for the next generation of power devices. This is primarily attributed to its faster switching speed, heightened thermal conductivity, and superior breakdown strength, which surpass the capabilities of conventional silicon-based power devices. GaN-based HEMTs have consistently demonstrated their excellence in radar, military, and telecommunications applications [11]. Furthermore, the polarization effect in GaN HEMTs generates a two-dimensional electron gas (2DEG) at the interface, thereby significantly augmenting electron mobility. This characteristic renders GaN-based devices exceptionally well suited for high-power applications. It is noteworthy to mention that AlGaN/GaN high-electron-mobility transistors (HEMTs) have previously shown their efficacy in high-power applications. The objective is to extend the operational frequency range of AlGaN/GaN high-electron-mobility transistors (HEMTs) in order to facilitate their utilization in millimeter-wave applications. Reducing the gate length of the device to the nanoscale range has significant importance. Many researchers have published their work on different gate geometry to enhance RF performance such as camel-shaped [12], T-shaped [13], and gamma-shaped [14] gate structures. Among all these gate structures, the T-shaped gate geometry is appropriate to enhance the RF

✉ Rishu Chaujar
chaujar.rishu@dtu.ac.in

[1] Department of Applied Physics, Delhi Technological University, New Delhi, Delhi, India

Published online: 05 February 2024

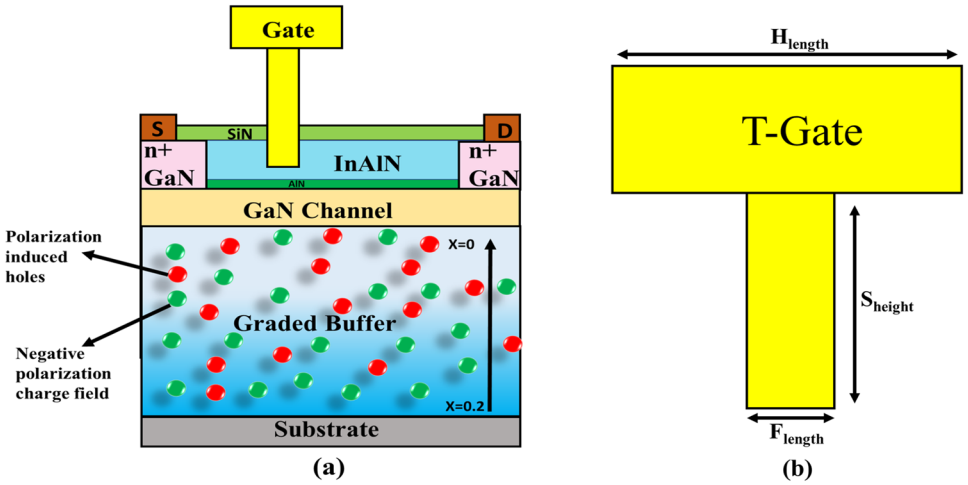


Fig. 1 **a** Proposed T-gate InAlN/AlN/ graded buffer HEMT, **b** T-gate structure

performance of the device. However, the DC and RF performance of a device is quite sensitive to the T-gate size. When the foot length of the T-gate is decreased to below 50 nm, issues such as high leakage current [15], decreased breakdown voltage [16], high current collapse [17], and parasitic resistance [18] have surfaced. The DC and RF properties of the device might be adversely affected by these problems. Inefficient power dissipation and shortened device lifetimes might result from excessive leakage current. High current collapse may cause instability and a gradual decline in device performance, while a low breakdown voltage might cause device failure. Parasitic resistance, on the other hand, can limit the device's maximum frequency of operation, leading to reduced RF performance. Therefore, selecting the proper T-gate size is crucial to enhancing the device's RF performance. Improper T-gate size selection can never improve the device frequency parameters, and it may even worsen the performance of the device.

Moreover, gate engineering is insufficient to suppress the short-channel effect (SCE). Some more buffer and barrier engineering is required to reduce the SCE and enhance the RF performance of the device. Hence, to enhance the performance of the T-gate HEMT device, some barrier and buffer engineering has been carried out in the proposed structure. The conventional AlGaN barrier layer is replaced by the lattice-matched InAlN material with a lattice constant of 0.83 [19]. The implementation of constructive buffer engineering involves the substitution of Fe doping [20] with polarization-induced doping in the buffer. The introduction of either iron (Fe) or carbon (C) doping in the buffer zone has been shown to improve the resistivity. However, this doping also leads to the formation of deep-level acceptors, which exacerbates the phenomenon of current collapse [21]. Polarization doping within the buffer layer serves as a barrier that impedes 2DEG in the buffer region. This, in turn, significantly improves the breakdown voltage and minimizes the leakage current. Many theoretical and experimental papers have shown the advantages of graded engineering. The impact of polarization bulk doping is demonstrated by Jena et al. [22]. The experimental work is presented by Li et al. [23] on polarization-doped pn junction. Polarization doping for U.V. LEDs is demonstrated by Simon et al. [24]. The combined incorporation of graded and T-gate engineering into the proposed structure improves the device's RF performance and makes it well suited for millimeter-wave applications.

This paper aims to deeply optimize the T-gate geometry to achieve the operating frequency of the mm-wave spectrum. The parasitic capacitance of the device directly affects the RF performance of the device. Hence, a proper study of parasitic capacitance is required. The impact of gate head (H_{length}), gate stem height (S_{height}), and gate foot length (F_{length}) on parasitic capacitance is deeply investigated in this paper. This paper is arranged as follows: Sect. 2 describes the device structure and simulation models, Sect. 3 presents the device fabrication feasibility, and Sect. 4 investigates the effect of gate geometry on parasitic capacitance, RF, and microwave performance of the device with variable gate head, stem height, and foot length. The paper is concluded in Sect. 5.

2 Design and Simulation Model for the Device

Figure 1a shows the two-dimensional structure of the proposed T-gate InAlN/AlN GaN HEMT with a graded buffer. Table 1 summarizes the design parameters. Within the proposed structure, an AlGaN-graded buffer layer has been implemented, characterized by a continuous alteration in Al composition along the (0001) direction, transitioning from 0.2 to 0. This linear change of composition formed the negative charge field in the buffer. Due to charge neutrality, a

Table 1 Device parameters

Parameters	Values
Passivation layer	20 nm
Barrier layer	6 nm
Gate to source length	400 nm
Channel layer	20 nm
Buffer layer	1 μm
Source /drain doping	1×10^{19} cm^{-3}
Gate to drain length	590 nm
Spacer layer	1 nm

substantial population of holes is generated by the negative charge field. This phenomenon plays a pivotal role in augmenting the breakdown voltage and concurrently diminishing the leakage current. The dimensions of the T-gate, i.e., gate head length, gate stem height, and gate foot length, were taken as variables to study the impact of gate geometry on device performance. For the simulation of the proposed structure, we employed the Silvaco TCAD software [25]. Models used in the simulation process are expressed in Table 2 [26].

3 Experimental Calibration, Fabrication Feasibility, and Simulation Methodology

Calibration of the physical models to the experimental paper [27] is shown in Fig. 2. There is remarkable congruence between the simulated and experimental outcomes. Figure 3a shows the fabrication procedure for the proposed structure. The total fabrication process for the proposed structure involves two approaches, i.e., the bottom-up and top-down approaches. The gate patterning and source/drain mesa are through a top-down approach. The growth of the epitaxial stack is carried out by the molecule beam epitaxy (MBE) [22]. The ohmic source/drain n + + GaN regions are regrown by MBE with SiO_2 growth mask [28]. The SiN passivation layer is placed on top of the epitaxial stack by plasma-enhanced chemical vapor deposition (PECVD) [29]. The fabrication of the T-gate consists following steps. The electron beam lithography is used for gate definition; gate foot is partially recessed into the barrier layer by dry etching and then the deposition of gate metal by an electron beam evaporator (EBE) [29]. Figure 3b illustrates the simulation algorithm for GaN HEMT transistors conducted within the Silvaco TCAD environment, utilizing the "Atlas" application. We utilized experimental data to simulate these devices, along with their operational characteristics.

Table 2 Physical models [21]

Models	Formula used	Explanation
High-field mobility	$\mu(E) = \frac{\mu(N,T)+V_{\text{sat}}\frac{E(N_1-1)}{E_C^{N_1}}}{1+a_n\left(\frac{E}{E_C}\right)^{N_2}+\left(\frac{E}{E_C}\right)^{N_1}}$	The electric field accelerates both holes and electrons, causing them to lose momentum and undergo scattering processes. Consequently, the scattering mechanism of the device has been assessed using mobility models
Low-field mobility	$\frac{1}{\mu(T,N)} = a\left(\frac{N}{N^{\text{ref}}}\right)\left(\frac{T}{300}\right)^{-15} \times \ln\left[1+3\left(\frac{T}{300}\right)^{2}\left(\frac{N}{N^{\text{ref}}}\right)^{\frac{-2}{3}}\right] + b\left(\frac{T}{300}\right)^{1.5} + \frac{c}{\left(e^{\frac{1065}{T}}\right)-1}$	For low-field mobility, Albrecht model is employed
Polarization model [26]	$P_{\text{total}} = [(P_{\text{PE}} + P_{\text{SP.}})(\text{bottom})] - [(P_{\text{PE}} + P_{\text{SP.}})(\text{top})]$	The III-nitrides are inherently pyroelectric, demonstrating both piezoelectric (P_{PE}) and spontaneous (P_{SP}) polarization
Recombination model	$R^{\text{SRH}} = \frac{pn-n_i^2}{\tau_p\left[n+n_i\exp\left(\frac{E_{\text{trap}}}{kT}\right)\right]+\tau_n\left[p+n_i\exp\left(\frac{-E_{\text{trap}}}{kT}\right)\right]}$	Shockley–Read–Hall (SRH) analysis explores the electron trapping mechanism at various trap levels

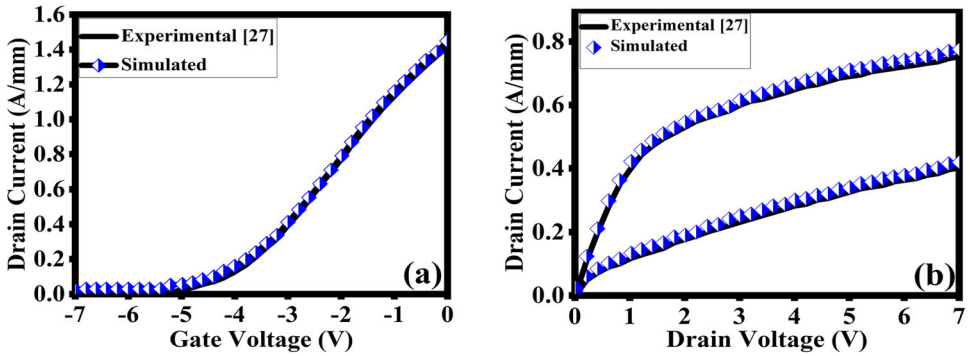
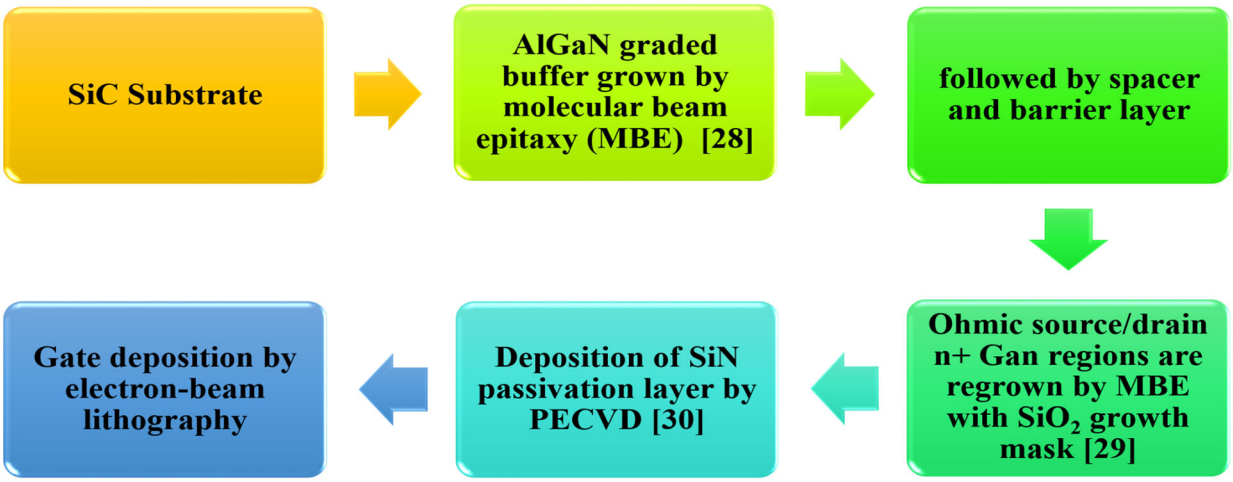
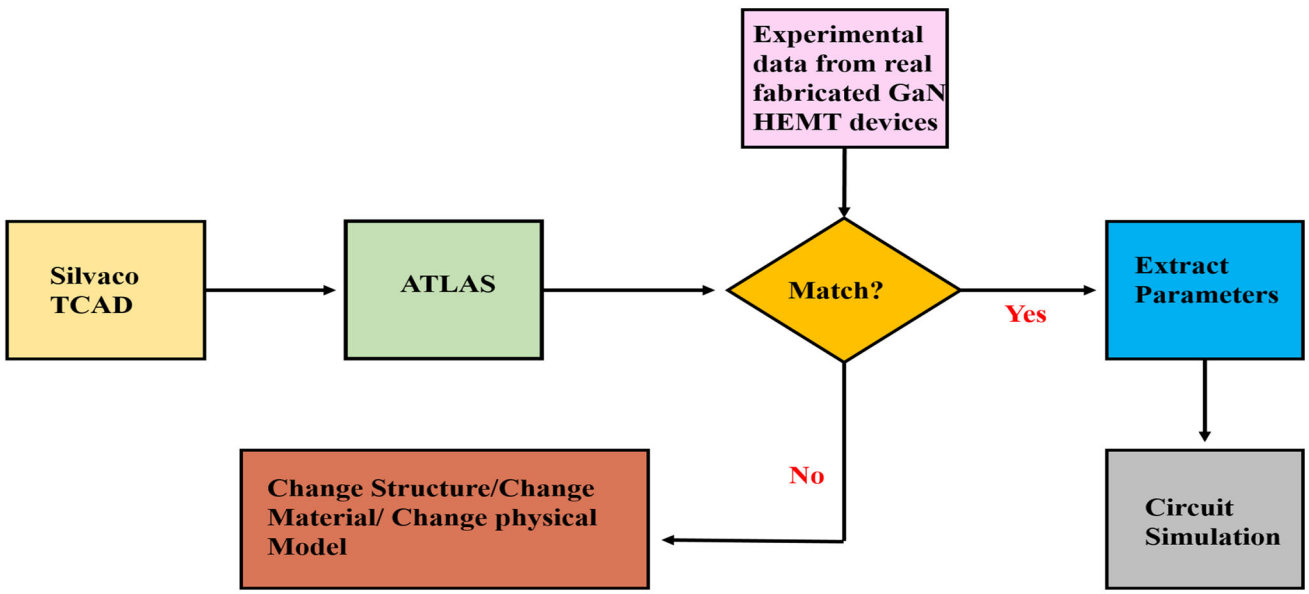


Fig. 2 **a** Transfer, **b** I_d–V_d curve of simulated and experimental data



(a) Fabrication flow of proposed device.



(b) Simulation methodology.

Fig. 3 **a** Fabrication flow of proposed device. **b** Simulation methodology

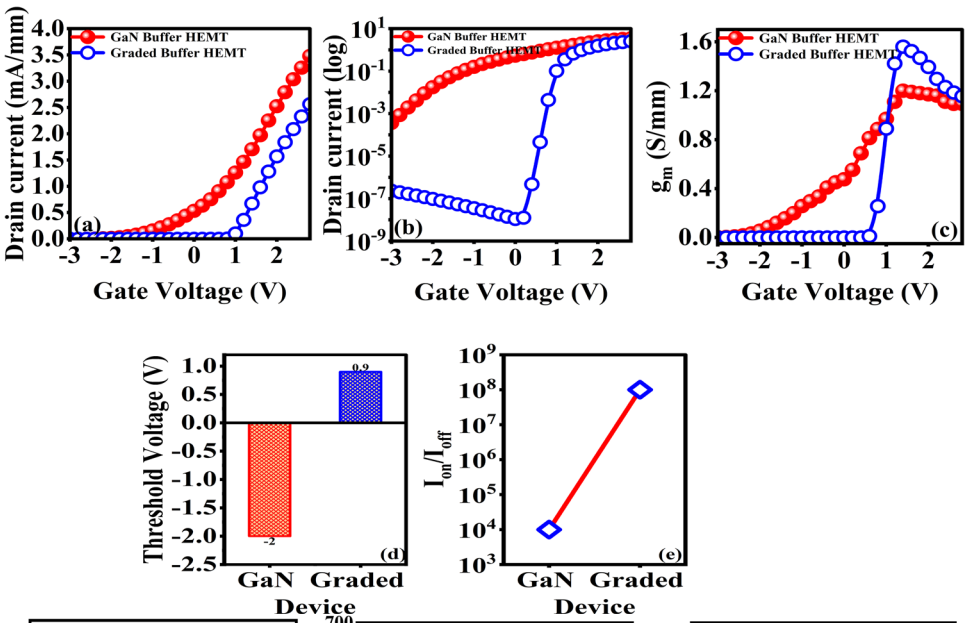


Fig. 4 Drain current characteristics of conventional and proposed HEMT **a** linear, **b** log scale, **c** transconductance characteristics, **d** threshold voltage, **e** I_{on}/I_{off} ratio

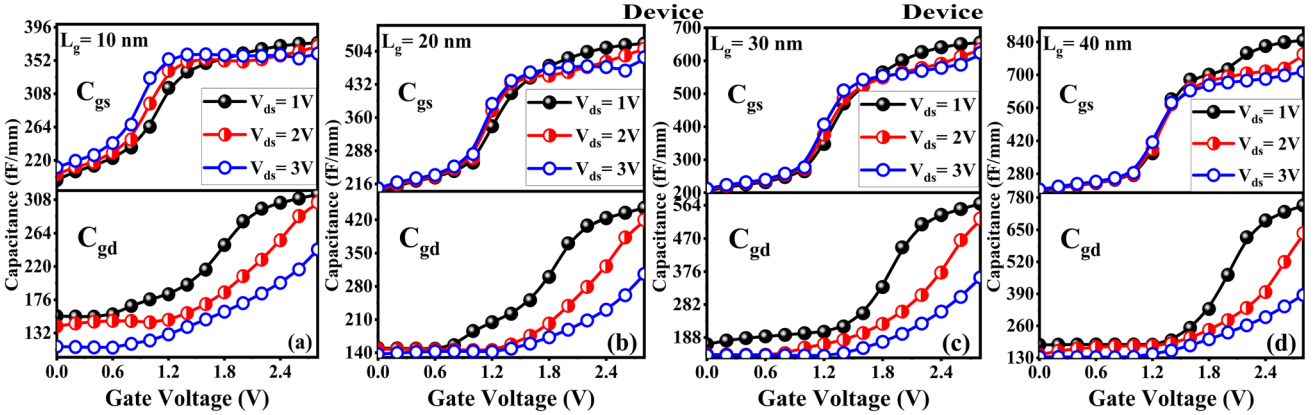


Fig. 5 Parasitic capacitance of proposed device for **a** F_{length} 10 nm, **b** F_{length} 20 nm, **c** F_{length} 30 nm, **d** F_{length} 40 nm for V_{ds} 1–3 V

Table 3 Comparison of device performance

Device	I_d (A/mm)	g_m (S/mm)	V_{th} (V)	I_{on}/I_{off}
GaN buffer (conventional)	3.48	1.26	− 2	10^4
Graded buffer HEMT (proposed)	2.56	1.55	0.9	10^8
%	35% ↓	23% ↑	145% ↑	99% ↑

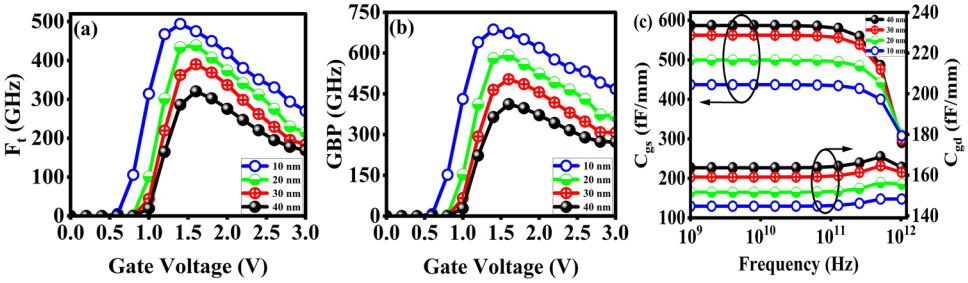


Fig. 6 Comparison of **a** F_t **b** GBP **c** C_{gs} and C_{gd} for variable F_{length}

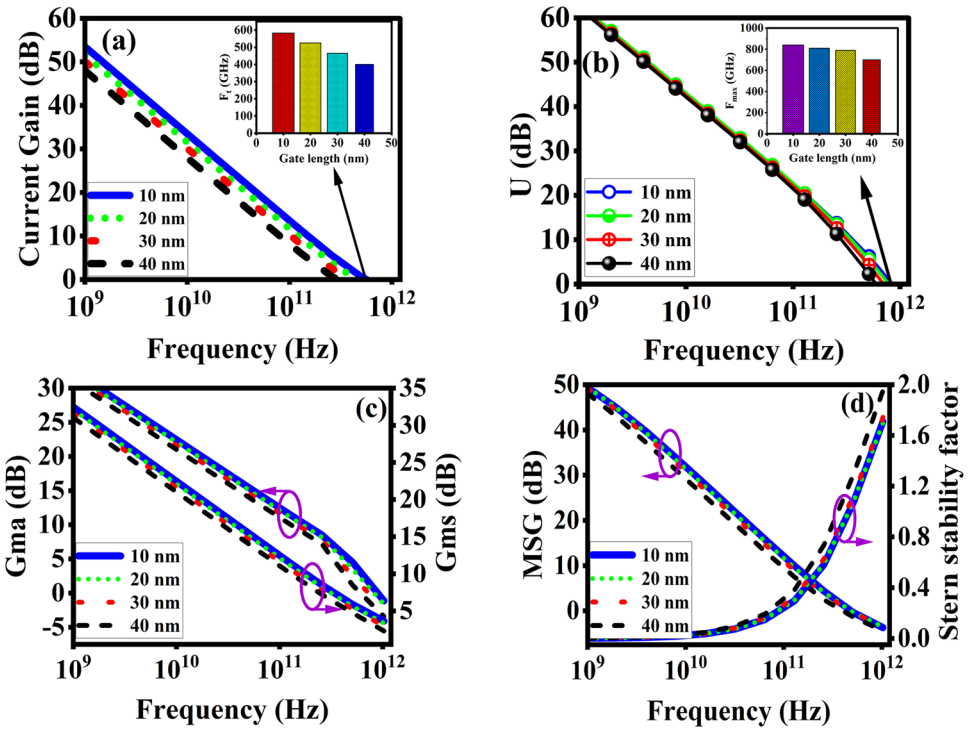


Fig. 7 Variation of **a** h_{21}, **b** U, **c** Gms and Gma, **d** MSG and k factor for different F_{length}

4 Results and Discussion

4.1 DC Characteristics

The transfer and electrical performance comparison of proposed (Graded buffer-HEMT) and conventional (GaN buffer-HEMT) devices is depicted in Fig. 4a–d. The lower drain current and leakage current are observed in Fig. 4a, b for a proposed device, further enhancing the $I_{\text{on}}/I_{\text{off}}$ ratio for graded buffer HEMT, as shown in Fig. 4e. The polarization-doped buffer is the main reason for reducing leakage current. The 2DEG confinement and gate controllability are improved by the presence of a negative polarization field in the graded buffer region, which further shifts the threshold voltage in a positive direction (0.9 V) compared to GaN HEMT, as depicted in Fig. 4d. Moreover, a higher transconductance peak is observed for the proposed device (1.55 S/mm), which improves by about 23% compared with the conventional ones, as depicted in Fig. 4c. The sharp change in transconductance at the pinch-off region is observed for the graded buffer HEMT, which exhibits the lower subthreshold current in the device. The parameters comparison of simulated and published results is tabulated in Table 3.

4.2 *Effect of Gate Foot Length (F_{length})*

This subsection investigates the impact of the gate foot length of the proposed structure on the parasitic capacitance, RF, and microwave characteristics of the device. Figure 5a–d exhibits the parasitic capacitance (C_{gs} and C_{gd}) dependency on F_{length} for drain voltage 1–3 V as a function of gate bias. Figure 5a–d shows that as the F_{length} reduced from 40 to 10 nm, the gate-to-source capacitance reduced to 54%, while the gate-to-drain capacitance reduced to 18%. This reduction in parasitic capacitance with a reduction in F_{length} is due to the redistribution of an electric field at the edges of the gate foot. Furthermore, with increasing the drain voltage from 1 to 3 V, parasitic capacitance starts decreasing. A large change in the C_{gd} is observed compared to the gate-to-source capacitance. This is because an increase in drain voltage leads to a decrease in the depletion layer near the drain end. The variation in cutoff frequency (F_{t}) for different F_{length} is expressed in Fig. 6a, and the mathematical expression for the F_{t} is expressed in Eq. (1) [30]. The reduction in F_{length} from 40 to 10 nm enhances the F_{t} by about 72%. This enhancement in F_{t} value for lower F_{length} is due to the decrement of parasitic capacitance with F_{length}. This enhancement in cutoff frequency makes the proposed device suitable for millimeter-wave applications. Another capacitance-dependent parameter is the gain–bandwidth product (GBP), which measures the amplifier's gain and bandwidth. The maximal gain of the device is represented by the peak value of GBP at a specific frequency. The mathematical expression for the GBP is shown in Eq. (2) [31]. The variability in GBP against gate bias is shown for various F_{length} in Fig. 6b. The value of GBP is seen to increase with gate bias at first, reaching a maximum at $V_{\text{gs}} = 1.5$ V and then declining with increasing gate bias. As shown in

Table 4 Comparison of device performance for variable gate foot length

F_{length} (nm)	C_{gs} (fF/mm)	C_{gd} (fF/mm)	F_{t} (GHz)	GBP (GHz)	F_{max} (GHz)
10	350	140	583	686	840
20	427	153	525	580	810
30	484	161	465	464	790
40	539	166	400	364	700

Fig. 5, this decrease is primarily attributable to the increase in capacitance at increasing gate bias. Furthermore, a higher value of GBP is reported for $F_{\text{length}} = 10$ nm, which is 88% greater than $F_{\text{length}} = 40$ nm. The increase in GBP value with decreasing gate length is due to the reduction in parasitic capacitance with gate length.

Capacitance versus frequency characteristics for various gate length are displayed in Fig. 6c. In the low-frequency range, it was noted that the capacitance was almost unaffected by the frequency; on the other hand, a decrease in capacitance was observed for the high-frequency range. Interface trap charges are present in the heterostructure, and they are responsible for the capturing and releasing of electrons, as well as the generation of interface trap capacitance at lower frequencies. Because of this, when the frequency is high, they no longer respond to the AC signal, and the interface trap capacitance becomes zero; this results in a significant decrease in capacitance. This further proved that the reduction in F_{length} reduces the parasitic capacitance. Thus, the best results are obtained at $F_{\text{length}} = 10$ nm.

$$F_{\text{t}} \approx g_m/2\pi(C_{\text{gs}} + C_{gd}) \tag{1}$$

$$\text{GBP} = \frac{g_m}{2\pi C_{\text{gs}}} \tag{2}$$

The microwave properties of graded buffer HEMT for variable F_{length} are shown in Fig. 7a–d. These analyses are also known as small-signal analysis as they require a small input signal for the characterization. Equations (3–4) expresses the mathematical expression of current gain (h_{21}) and unilateral power gain (U) in terms of S-parameters [32]. Figure 7a, b exhibits the variation in h_{21} and U with frequency for different F_{length}.

$$h_{21} = \left| \frac{-2S_{21}}{(1 - S_{11}) \times (2 - S_{22}) - S_{12} \times S_{21}} \right| \tag{3}$$

$$U = \left(\frac{|S_{21}|^2}{\left(1 - |S_{11}|^2\right) \times \left(1 - |S_{22}|^2\right)} \right) \tag{4}$$

At unity current gain and unity unilateral power gain, the device's F_{t} and F_{max} are determined. Figure 7a, b has an inset bar graph showing F_{t} and F_{max} at varying F_{lengths}. The higher value of F_{t} (523 GHz) and F_{max} (840 GHz) is observed for

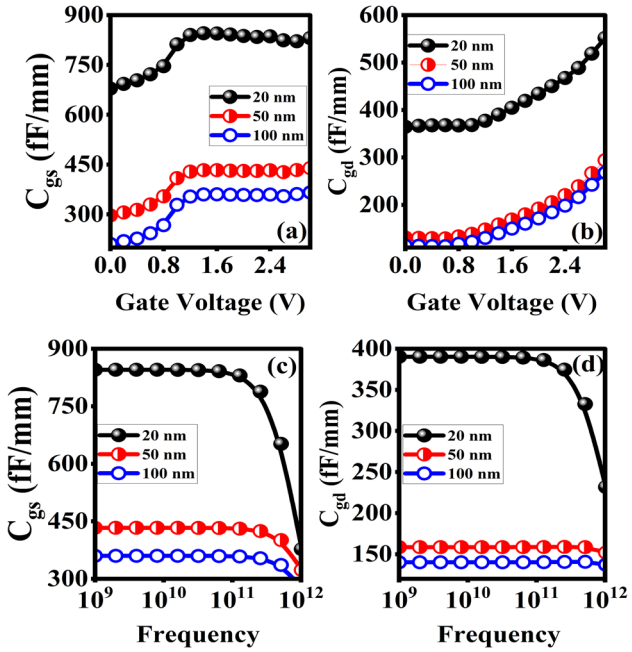


Fig. 8 Impact of S_{height} on **a** C_{gs}, **b** C_{gd} with gate bias, **c** C_{gs}, **d** C_{gd} with frequency

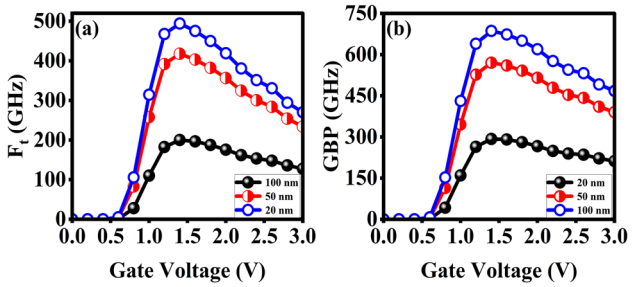


Fig. 9 Impact of S_{height} on **a** $F_{\text{t,}}$ **b** GBT with gate bias

the $F_{\text{length}} = 10$ nm, which is 45% and 20% higher than the $F_{\text{length}} = 40$ nm. This enhancement is basically due to the reduction in parasitic capacitance (C_{gs} and C_{gd}) with gate length, as shown in Fig. 6d. The reduction in parasitic capacitance with the reduction in F_{length} is due to the redistribution of an electric field at the edges of the gate foot.

Figure 7c displays the relationship between F_{length} and the maximum stable and available power gain Gms and Gma, respectively, of the device. Equation (5) [33] is used to define

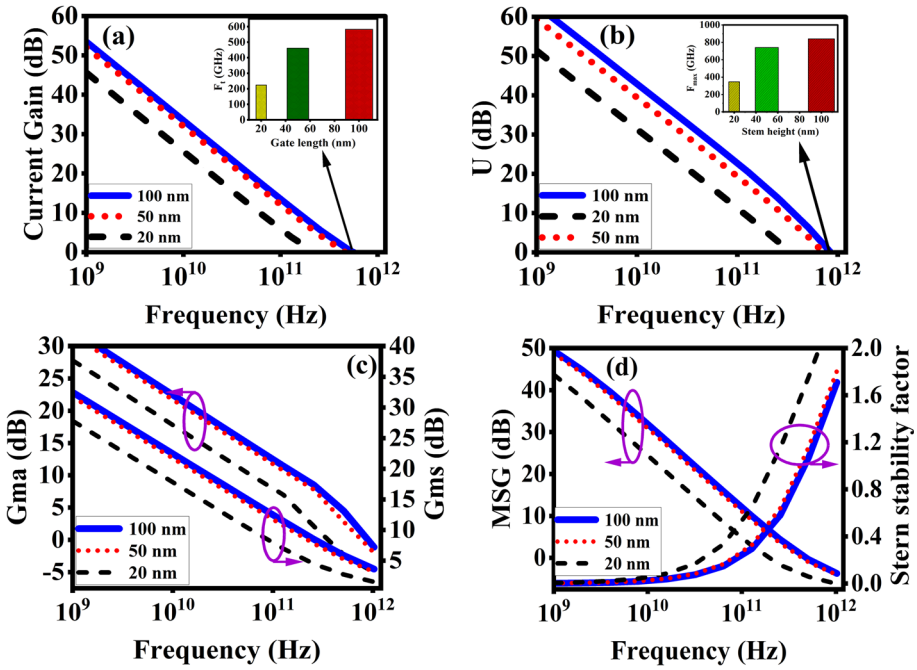


Fig. 10 Variation in **a** h_{21}, **b** U, **c** Gms and Gma, **d** MSG and k factor for different S_{height}

Table 5 Comparison of device performance for variable gate stem height

S_{height} (nm)	C_{gs} (fF/mm)	C_{gd} (fF/mm)	F_{t} (GHz)	GBP (GHz)	F_{max} (GHz)
20	788	374	224	292	346
50	424	158	460	570	741
100	350	140	583	686	840

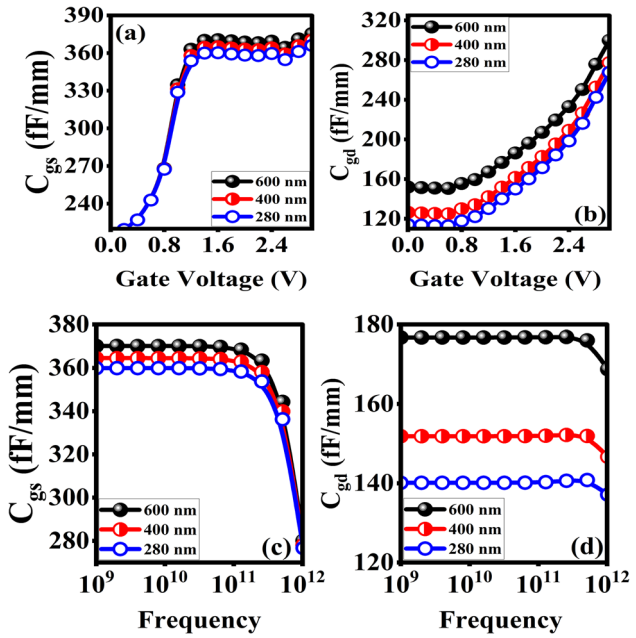


Fig. 11 Impact of H_{length} on **a** C_{gs}, **b** C_{gd} with gate bias, **c** C_{gs}, **d** C_{gd} with frequency

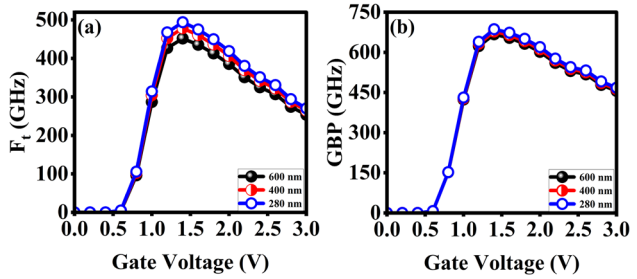


Fig. 12 Impact of H_{length} on **a** F_{t} **b** GBT with gate bias

the expression of Gma. The study of Gma and Gms is crucial for RF amplifier design. Figure 7c depicts that scaling F_{length} to 10 nm enhances the Gma and Gms by 50% and 40%. The Rollett stability factor ($K > 1$) serves as a crucial determinant of the device's stability, and its expression is defined by Eq. (6) [34]. Once the stipulated stability condition is met, the calculation of the MSG becomes possible, utilizing the k factor and S parameters, as demonstrated in Eq. (7).

Figure 7d illustrates the stability factor and MSG for various F_{length} values. Notably, a remarkable 20% improvement in MSG is evident with a reduction in F_{length} to 10 nm. Conversely, a high stability factor is observed in the case of 40 nm

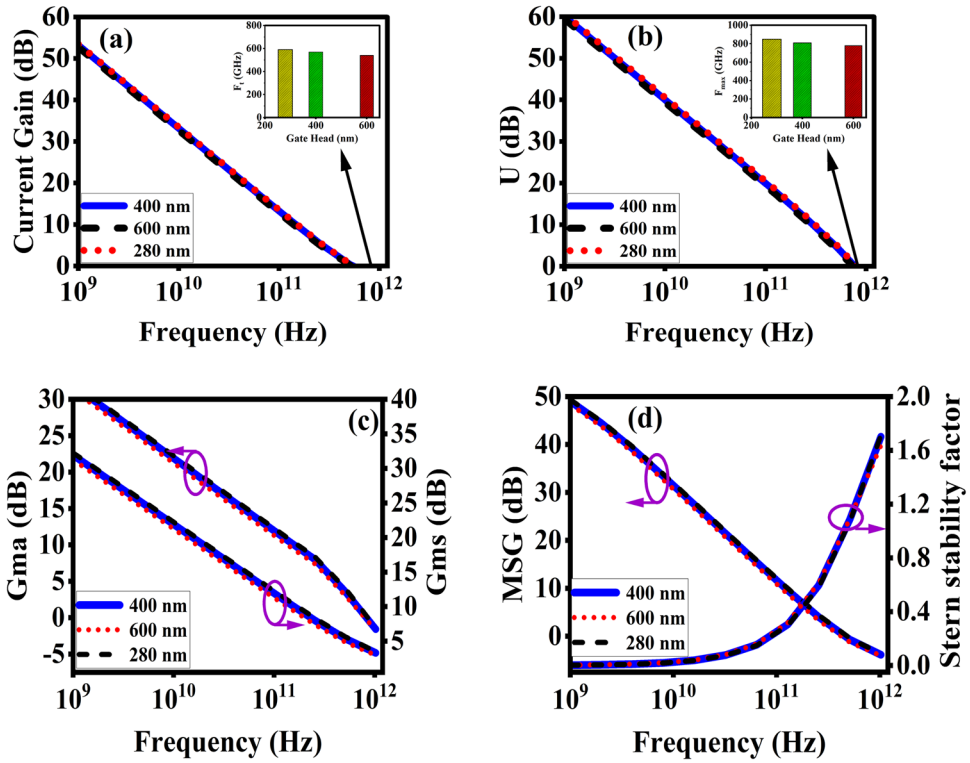


Fig. 13 Variation in **a** h_{21}, **b** U, **c** Gms and Gma, **d** MSG and k factor for different H_{length}

Table 6 Comparison of device performance for variable gate head length

H_{length} (nm)	C_{gs} (fF/mm)	C_{gd} (fF/mm)	F_{t} (GHz)	GBP (GHz)	F_{max} (GHz)
280	350	140	583	686	840
400	358	152	562	678	814
600	363	176	532	668	781

Table 7 Comparison of proposed structure with experimental and simulated published papers on T-gate hemt

Reference	L_{g}	I_{d} (A/mm)	C_{gs} (fF/mm)	C_{gd} (fF/mm)	F_{t} (GHz)	F_{max} (GHz)
[35]	20 nm	2.6	258	511	343	236
[13]	30 nm	2.1	–	–	350	340
[36]	30 nm	2.4	434	173	246	290
[37]	10 nm	2.5	315	121	524	758
[38]	30 nm	2.7	572	84	426	366
[39]	30 nm	0.6	–	–	50	150
[40]	50 nm	2.8	444	104	145	218
This work	10 nm	2.5	350	140	583	840

F_{length}. It is noteworthy that for 10 nm F_{length}, the stability factor exhibits $K > 1$ for high frequencies and $K < 1$ for low frequencies, rendering it a suitable choice for amplifier design. The impact of F_{length} on device performance is tabulated in Table 4.

$$\text{Gma} = \frac{P_{\text{load max}}}{P_{\text{source max}}} \tag{5}$$

$$K = \frac{1 - |S_{21}|^2 - |S_{12}|^2 + |S_{11}S_{22} - S_{12}S_{21}|^2}{2 \times |S_{12}|^2 \times |S_{12}|^2} \tag{6}$$

$$\text{MSG} = \frac{|S_{21}|^2}{|S_{12}|^2} \times \left(K \pm \sqrt{(K^2 - 1)}\right) \tag{7}$$

4.3 *Effect of Gate Stem Height (S_{height})*

This subsection investigates the impact of the gate stem height of the proposed structure on the parasitic capacitance, RF, and microwave characteristics of the device. Figure 8a–d investigates the impact of S_{height} on the parasitic capacitances. It has been observed that as the stem height increases from 20 to 100 nm, the parasitic capacitance decreases. The increase in the gate stem height of the device causes an increase in the distance between the gate source and the gate drain, which results in a decrease in the parasitic capacitance of the device. The graph of parasitic capacitance versus gate voltage is shown in Fig. 8a, b, and it depicts the values for S_{height} 20, 50, and 100 nm. The 125% and 167% reduction in C_{gs} and C_{gd} is observed for increasing the S_{height} from 20 to 100 nm. Moreover, the variation in parasitic capacitance with frequency is shown in Fig. 8c, d. This further proved that the increment of S_{height} reduces the parasitic capacitance. Figure 9a, b illustrates how the cutoff frequency and gain–bandwidth product change in response to changes in gate voltage. The increment in S_{height} to 100 nm enhances the F_{t} and GBP by about 147% and 132%. This improvement in F_{t} and GBP parameters is due to the significant reduction in parasitic capacitance.

The impact of S_{height} on microwave performance is shown in Fig. 10a–d. The impact of S_{height} on current gain and unilateral gain is depicted in Fig. 10a, b. The calculated value of F_{t} and F_{max} at which current gain and unilateral gain become unity is portrayed by the inset bar graph of Fig. 10a, d. The 160% and 142% enhancement of F_{t} and F_{max} is observed when S_{height} increases to 100 nm. In Fig. 10c, the 304% and 88% enhancement in Gma and Gms is observed on increasing the S_{height} from 20 to 100 nm. Increment in S_{height} further improves the maximum stable gain by about 239%. The stability factor for S_{height} 100 nm in the cutoff frequency range is 1.08, which shows the stable frequency operation in the device. Table 5 summarizes the effects of S_{height} on device performance. Thus, the best results are obtained at $S_{\text{height}} =$ 100 nm.

4.4 *Effect of Gate Head Length (H_{length})*

This subsection investigates the impact of the gate head length of the proposed structure on the parasitic capacitance, RF, and microwave properties of the device. Figure 11a–d investigates the impact of H_{length} on the parasitic capacitances. A parallel increase in parasitic capacitance has been seen as H_{length} is reduced from 600 to 280 nm. This is because the increase in H_{length} regulates the depletion layer by providing a consistent electric field distribution below the gate edges. Because of the redistribution of the electric field at the gate edges, the channel depletion length is suppressed and lengthened, which consequently increases capacitance.

Figure 11a, b exhibits the graph of parasitic capacitance with gate voltage for H_{length} 280, 400, and 600 nm. It has been observed that C_{gs} and C_{gd} increase gradually at low gate voltage but rapidly at high gate bias due to an increase in the lateral electric field that speeds up the movement of carriers from the drain to the source. The 2.8% and 25% reduction in C_{gs} and C_{gd} is observed for reducing the H_{length} from 600 to 280 nm. Furthermore, the variation in parasitic capacitance with frequency is shown in Fig. 11c, d. Figure 12a, b illustrates how the cutoff frequency and the gain–bandwidth product change as a function of the gate bias. The decrement in H_{length} to 280 nm enhances the F_{t} and GBP by about 9.5% and 2.5% due to the significant reduction in parasitic capacitance with H_{length}. The impact of H_{length} on microwave performance is shown in Fig. 13a–d. The less impact of H_{length} on h_{21} and unilateral gain is observed in Fig. 13a, b. The calculated value of F_{t} and F_{max} at which current gain and unilateral gain become unity is portrayed by the inset bar graph of Fig. 13a, b. The 9.5% and 7.5% enhancement of F_{t} and F_{max} is observed when H_{length} decreases to 280 nm. In Fig. 13c, a small enhancement in Gma and Gms is observed on increasing the H_{length} from 600 to 280 nm. A minor improvement is observed in maximum stable gain and stability factor for different H_{length} shown in Fig. 13d. The impact of H_{length} on device performance is tabulated in Table 6. Table 7 shows the comparison between simulated and published results. Thus, the best results are obtained at $H_{\text{length}} = 280$ nm.

5 Conclusion

This work explored the proper design of T-gate InAlN/AlN/graded buffer HEMT and investigated their effects on the device's intrinsic characteristics, particularly focusing on parasitic capacitance, radio-frequency (RF), and microwave performance. The simulation results shed light on the impact of various design parameters. In this study, we have focused on specifically three critical design parameters:

- H_{length}: The reduction in H_{length} from 600 to 280 nm significantly reduced the C_{gs} by 2.8% and C_{gd} by 25% which significantly enhance the F_{t} and F_{max} by about 9.5% and 7.5% due to the significant reduction in parasitic capacitance with H_{length}.
- F_{length}: The reduction in F_{length} from 40 to 10 nm significantly reduced the C_{gs} by 54% and C_{gd} by 18% which significantly enhance the F_{t} and F_{max} by about 45% and 20% due to the significant reduction in parasitic capacitance with F_{length}.
- S_{height}: The increment in S_{height} from 20 to 100 nm significantly reduced the C_{gs} by 125% and C_{gd} by 167% which

significantly enhance the F_t and F_{max} by about 160% and 142% due to the significant reduction in parasitic capacitance with increment in S_{height}.

The proper design of T-gate InAlN/AlN/graded buffer HEMT devices, with particular attention to gate dimensions, has yielded encouraging results. The scaling of H_{length}, F_{length} adjustments, and an increment in S_{height} have collectively led to a substantial reduction in parasitic capacitance, paving the way for enhanced RF and microwave performance, making these devices more suitable for high-frequency applications.

Author Contributions All authors contributed to the study's conception and design.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability The authors mentioned above have all the relevant data associated with this research work and will be dedicated to sharing that they will be asked to do so in the future.

Declarations

Conflict of interest The authors declare that they have no known conflict of interest or personal relationships that could have appeared to influence the work reported in this paper.

Compliance with Ethical Standards The authors have seen all the Ethical Standards and will be supposed to follow them in the future.

Consent to Participate and for Publication Since the concerned research paper is based on the 'non-life science journal.' So, it is 'not applicable' here. However, the authors have gone through all journal policies and consented to the authorities for further processing.

References

1. Rajawat, V.S.; Kumar, A.; Choudhary, B.: Impact on DC and analog/RF performances of SOI based GaN FinFET considering high-k gate oxide. Mem. Mater. Devices Circuits Syst. **65**(3), 860–866 (2018). https://doi.org/10.1109/TED.2018.2793853
2. Rao, M.V.G.; Ramanjaneyulu, N.; Pydi, B.; Soma, U.; Babu, K.R.; Prasad, S.H.: Enhancing performance of dual-gate FinFET with high-K gate dielectric materials in 5 nm technology: a simulation study. Trans. Electr. Electron. Mater. (2023). https://doi.org/10.1007/s42341-023-00473-5
3. Maurya, R.K.; Bhowmick, B.: Review of FinFET devices and perspective on circuit design challenges. SILICON **14**, 5783–5791 (2022). https://doi.org/10.1007/s12633-021-01366-z
4. Kumar, A.; Tripathi, M.M.; Chaujar, R.: Comprehensive analysis of sub-20 nm black phosphorus based junctionless-recessed channel MOSFET for analog/RF applications. Superlattices Microstruct. **116**, 171–180 (2018). https://doi.org/10.1016/j.spmi.2018.02.018
5. Kumar, A.: Effect of trench depth and gate length shrinking assessment on the analog and linearity performance of TGRC-MOSFET. Superlattices Microstruct. **109**, 626–640 (2017). https://doi.org/10.1016/j.spmi.2017.05.045
6. Kumar, A.; Tripathi, M.M.; Chaujar, R.: Reliability issues of In_2O_5Sn gate electrode recessed channel MOSFET: impact of interface trap charges and temperature. IEEE Trans. Electron Devices **65**(3), 860–866 (2018). https://doi.org/10.1109/TED.2018.2793853
7. Dutt, A.; Tiwari, S.; Upadhyay, A.K.; Mathew, R.; Beohar, A.: Impact of drain underlap and high bandgap strip on cylindrical gate all around tunnel FET and its influence on analog/RF performance. SILICON **14**, 9789–9796 (2022). https://doi.org/10.1007/s12633-022-01692-w
8. Vimala, P.; Arun Samuel, T.S.: Investigation of cylindrical channel gate all around InGaAs/InP heterojunction heterodielectric tunnel FETs. SILICON **13**, 3899–3907 (2021). https://doi.org/10.1007/s12633-020-00691-z
9. Singh, M.; Chaudhary, T.; Raj, B.: Design and simulation analysis of silicon germanium nanowire FET for low power applications. SILICON **15**, 1491–1497 (2023). https://doi.org/10.1007/s12633-022-02101-y
10. Singh, S.; Solay, L.R.; Anand, S.; Kumar, N.; Ranjan, R.; Singh, A.: Implementation of gate-all-around gate-engineered charge plasma nanowire FET-based common source amplifier. Micromachines **14**, 1357 (2023). https://doi.org/10.3390/mi14071357
11. Mishra, U.K.; Parikh, P.; Wu, Y.F.: AlGaN/GaN HEMTs-an overview of device operation and applications. Proc. IEEE **90**, 1022–1031 (2002). https://doi.org/10.1109/JPROC.2002.1021567
12. Khan, M.A.; Heo, J.W.; Kim, Y.J.; Park, H.C.; Park, H.M.; Kim, H.S.; Mun, J.K.: Effects of a recessed camel-gate head structure on normally-off ALGaN/GaN HEMTs. J. Korean Phys. Soc. **62**, 787–793 (2013). https://doi.org/10.3938/jkps.62.787
13. Murugapandiyan, P.; Ravimaran, S.; William, J.: 30 nm T-gate enhancement-mode InAlN/AlN/GaN HEMT on SiC substrates for future high power RF applications. J. Semicond. **38**, 084001 (2017). https://doi.org/10.1088/1674-4926/38/8/084001
14. Peroni, M.; Romanini, P.; Pantellini, A.; Mariucci, L.; Minotti, A.; Ghione, G.; Camarchia, V.; Limiti, E.; Serino, A.; Chini, A.: Design, fabrication, and characterization of gamma-gate GaN HEMT for high-frequency/wide-band applications. In: Proceedings of 31st Workshop on Compound Semiconductor Devices and Integrated Circuits (WOCSDICE), pp. 371–378 (2007)
15. Deng, J.; Shao, J.; Wan, J.; Lu, B.; Chen, Y.: A theoretical study of gating effect on InP-InGaAs HEMTs by tri-layer T–shape gate. Microelectron. Eng. **208**, 54–59 (2019). https://doi.org/10.1016/j.mee.2019.02.004
16. Murugapandiyan, P.; Ravimaran, S.; William, J.: DC and microwave characteristics of Lg 50nm T-gate InAlN/AlN/GaN HEMT for future high power RF applications. AEU-Int. J. Electron. C. **77**, 163–168 (2017). https://doi.org/10.1016/j.aeue.2017.05.004
17. Saadaoui, S.; Fathallah, O.; Maaref, H.: Effects of current transportation and deep traps on leakage current and capacitance hysteresis of AlGaN/GaN HEMT. Mater. Sci. Semicond. Process. **115**, 105100 (2020). https://doi.org/10.1016/j.mssp.2020.105100
18. Cao, X.; Thoms, S.; Macintyre, D.; McLelland, H.; Boyd, E.; Elgaid, K.; Hill, R.; Stanley, C.R.; Thayne, I.G.: Fabrication and performance of 50 nm T-gates for InP high electron mobility transistors. Microelectron. Eng. **73**, 818–821 (2004). https://doi.org/10.1016/j.mee.2004.03.058
19. Neuburger, M.; Zimmermann, T.; Kohn, E.; Dadgar, A.; Schulze, F.; Krtschil, A.; Gunther, M.; Witte, H.; Blasing, J.; Krost, A.; Daumiller, I.: Unstrained InAlN/GaN HEMT structure. In: Proceedings. IEEE Lester Eastman Conference on High-Performance Devices, vol.14, pp. 161–166 (2004) https://doi.org/10.1109/LECHPD.2004.1549688
20. Pampili, P.; Parbrook, P.J.: Doping of III-nitride materials. Mater. Sci. Semicond. Process. **62**, 180–191 (2017). https://doi.org/10.1016/j.mssp.2016.11.006

21. Uren, M.J.; Moreke, J.; Kuball, M.: Buffer design to minimize current collapse in GaN/AlGaN HFETs. IEEE Trans. Electron Devices **59**, 3327–3333 (2012). https://doi.org/10.1109/TED.2012.2216535
22. Jena, D.; Heikman, S.; Green, D.; Buttari, D.; Coffie, R.; Xing, H.; Keller, S.; DenBaars, S.; Speck, J.S.; Mishra, U.K.; Smorchkova, I.: Realization of wide electron slabs by polarization bulk doping in graded III–V nitride semiconductor alloys. Appl. Phys. Lett. **81**, 4395–4397 (2002). https://doi.org/10.1063/1.1526161
23. Li, S.; Ware, M.; Wu, J.; Minor, P.; Wang, Z.; Wu, Z.; Jiang, Y.; Salamo, G.J.: Polarization induced PN-junction without dopant in graded AlGaN coherently strained on GaN. Appl. Phys. Lett. **101**, 122103 (2012). https://doi.org/10.1063/1.4753993
24. Yao, C.; Yang, G.; Li, Y.; Sun, R.; Zhang, Q.; Wang, J.; Gao, S.M.: Investigation of N-face AlGaN ultraviolet light-emitting diodes with composition-varying AlGaN electron blocking layer. Opt. Quantum Electron. **48**, 1–9 (2015). https://doi.org/10.1007/s11082-015-0345-5
25. ATLAS User's Manual: SILVACO International. CA, USA (2016)
26. Ambacher, O.; Majewski, J.; Miskys, C.; Link, A.; Hermann, M.; Eickhoff, M.; Stutzmann, M.; Bernardini, F.; Fiorentini, V.; Tilak, V.; Schaff, B.: Pyroelectric properties of Al(In)GaN/GaN hetero- and quantum well structures. J. Phys. Condens. Matter **14**, 3399–3434 (2002). https://doi.org/10.1088/0953-8984/14/13/302
27. Han, T.; Dun, S.; Lü, Y.; Gu, G.; Song, X.; Wang, Y.; Xu, P.; Feng, Z.: 70-nm-gated InAlN/GaN HEMTs grown on SiC substrate with $f_T/f_{max} > 160$ GHz. J. Semicond. **37**, 4–8 (2016). https://doi.org/10.1088/1674-4926/37/2/024007
28. Guo, J.; Li, G.; Faria, F.; Cao, Y.; Wang, R.; Verma, J.; Gao, X.; Guo, S.; Beam, E.; Ketterson, A.; Schuette, M.: MBE-regrown ohmic in InAlN HEMTs with a regrowth interface resistance of 0.05 ˙Ωmm. IEEE Electron Device Lett. **33**, 525–527 (2012). https://doi.org/10.1109/LED.2012.2186116
29. Micovic, M.; Brown, D.F.; Kurdoghlian, A.; Santos, D.; Grabar, B.; Magadia, J.; Khalaf, I.; Tai, H.Y.; Prophet, E.; Burnham, S.D.; Wong, J.C.: GaN DHFETs having 48% power added efficiency and 57% drain efficiency at V-band. IEEE Electron Device Lett. **38**, 1708–1711 (2017). https://doi.org/10.1109/LED.2017.2763940
30. Geng, L.; Zhao, H.; Yu, K.; Ren, X.; Yang, D.; Song, Y.: Analysis of InGaN back-barrier on linearity and RF performance in a graded-channel HEMT. J. Electron. Mater. **52**, 1426–1436 (2023). https://doi.org/10.1007/s11664-022-10109-8
31. Goyal, P.; Kaur, H.: Impact of dual material gate design and retrograde channel doping on β-Ga_2O_3 MOSFET for high power and RF applications. SILICON **15**, 1597–1608 (2023). https://doi.org/10.1007/s12633-022-02079-7
32. Malik, P.; Gupta, R.S.; Chaujar, R.; Gupta, M.: AC analysis of nanoscale GME-TRC MOSFET for microwave and RF applications. Microelectron. Reliab. **52**(1), 151–158 (2012). https://doi.org/10.1016/j.microrel.2011.07.070
33. Bozorgi, S.A.; Orouji, A.A.: High performance nanoscale SOI MOSFET with enhanced gate control. Micro Nanostruct. **175**, 207523 (2023)
34. Kumar, A.; Gupta, N.; Goyal, A.K.; Massoud, Y.: Comprehensive power gain assessment of GaN-SOI-FinFET for improved RF/wireless performance using TCAD. Micromachines (Basel) **13**(9), 1418 (2022). https://doi.org/10.3390/mi13091418
35. Murugapandiyan, P.; Ravimaran, S.; William, J.; Ajayan, J.; Nirmal, D.: DC and microwave characteristics of 20 nm T-gate InAlN/GaN high electron mobility transistor for high power RF applications. Superlattices Microstruct. **109**, 725–734 (2017). https://doi.org/10.1016/j.spmi.2017.05.060
36. Murugapandiyan, P.; Ravimaran, S.; William, J.; Meenakshi Sundaram, K.: Design and analysis of 30 nm T-gate InAlN/GaN HEMT with AlGaN back-barrier for high power microwave applications. Superlattices Microstruct. **111**, 1050–1057 (2017). https://doi.org/10.1016/j.spmi.2017.08.002
37. Zine-eddine, T.; Zahra, H.; Zitouni, M.: Design and analysis of 10 nm T-gate enhancement-mode MOS-HEMT for high power microwave applications. J. Sci. Adv. Mater. Devices **4**, 180–187 (2019). https://doi.org/10.1016/j.jsamd.2019.01.001
38. Murugapandiyan, P.; Mohanbabu, A.; Lakshmi, V.R.; Ramakrishnan, V.N.; Varghese, A.; Wasim, M.O.H.D.; Baskaran, S.; Kumar, R.S.; Janakiraman, V.: Performance analysis of HfO2/InAlN/AlN/GaN HEMT with AlN buffer layer for high power microwave applications. J. Sci. Adv. Mater. Devices **5**, 192–198 (2020). https://doi.org/10.1016/j.jsamd.2020.04.007
39. Four, I.; Kameche, M.: Optimization of DC and AC performances for $Al_{0.26}Ga_{0.74}N$/GaN/4H-SiC. Int. J. Nanoelectron. Mater. **13**, 361–372 (2020)
40. Cui, P.; Zeng, Y.: Scaling behavior of InAlN/GaN HEMTs on silicon for RF applications. Sci. Rep. **12**, 16683 (2022). https://doi.org/10.1038/s41598-022-21092-9

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.

Indian Journal of Pure & Applied Physics
Vol. 62, February 2024, pp. 116-123
DOI: 10.56042/ijpap.v62i2.7700

Effect of Multi-step Sintering on the density of $Li_7La_3Zr_{1.75}Ce_{0.25}O_{12}$ as Solid State Electrolyte Material

Sharad Singh Jadaun[a], Amrish K Panwar[a*] & Geetanjali[b]

[a]Department of Applied Physics, Delhi Technological University, Delhi 110 042, India
[b]Department of Chemistry, KMC, University of Delhi, Delhi 110 007, India

Received 31 December 2023; accepted 22 January 2024

The development of garnets structured lithium lanthanum zirconate, $Li_7La_3Zr_2O_{12}$ (LLZO) have many amicable characteristics such as high stability against lithium metal anode and wide electrochemical voltage window that make them good solid electrolytic materials for lithium-ion batteries. However, the poor interfacial properties and limited ionic conductivity continue to limit their wide range of applications. Hence, the current study has been focused on improving their ionic conductivity as well as the dense microstructure. This study demonstrates a simple and efficient method to produce compact ceramics with high Li^+ conductivity using a multi-step sintering process and variation in grain growth, relative density with the temperature and time variation has been studied. Therefore, garnet-structured oxides with the nominal chemical formula: $Li_7La_3Zr_{1.75}Ce_{0.25}O_{12}$ (LLZCO) have been synthesized. TGA, XRD and SEM characterization has been performed to see optimum calcination temperature, phase formation and microstructural analysis of the sample, respectively. The average crystallite size of the prepared LLZCO is estimated about 100 nm as calculated by Scherrer's formula from intense peaks of XRD pattern. Further, the ionic conductivity, electronic conductivities, activation energies along with the electrochemical voltage window stability have been explored.

Keywords: Solid state electrolyte; Solid state battery; LLZO; Lanthanide doping

1 Introduction

Lithium-ion batteries (LIBs) have a significant impact on the market as energy storage devices in today's information-rich world and are utilized as energy-storing devices for a variety of applications starting from light to heavy duty applications. In contrast to previous battery systems, LIBs have a high gravimetric energy density, long cycle life, and high volumetric energy density[1-2]. However, thermal instability, high flammability, and electrochemical instability of liquid electrolytes do have certain negative effects on lithium-ion batteries performance[2-3]. Moreover, the performance of liquid electrolytes is also limiting for high voltage cathode and lithium metal anode[1-4], lithium metal anodes in liquid electrolytes cause battery degradation owing to dendrite formation[5]. To overcome these challenges, solid-state electrolyte may be the potential alternative. As the solid state electrolytes are stable against high voltage cathode and lithium metal anode. They provide wide electrochemical voltage window for high energy LIBs.

An ideal solid electrolyte should possess good electrochemical stability, high ionic conductivity with low electronic conductivity and good mechanical stability[6-7]. Solid electrolyte with a garnet structure is a viable option for LIBs. Thangadurai *et al*. has developed garnet type Li oxides $Li_5La_3Nb_2O_{12}$ and $Li_5La_3Ta_2O_{12}$, with the room temperature ionic conductivity of about 10^{-6} Scm^{-1}[ref 8]. Murugan *et al*.[9] synthesized the $Li_7La_3Zr_2O_{12}$(LLZO) with high conductive cubic phase garnet structure than the less conductive tetragonal phase was synthesized by Awaka *et al*.[10]. Li-ion occupancy is the basic difference between these two structures, in which tetragonal garnet has three Lithium sites which occupy tetrahedral site (8a), octahedral site (16f) and octahedral site (32g). In cubic LLZO, Li occupies two types of sites as (24d) tetrahedral site and (96h) distorted octahedral site. Hence, for this cubic LLZO structure, Li-Li ion bond length is short which results in high ionic conductivity[11]. Despite of having high mechanical strength and high electrochemical voltage window, there are some of the major challenges for practical applications such as; (a) high electrode electrolyte interfacial resistance, (b) higher sintering temperature (c) lower ionic conductivity then liquid electrolyte.

To address the ionic conductivity issue, the doping in garnets with aliovalent ions (Al, Nb, Ta) is the

*Corresponding author: (E-mail: amrish.phy@dtu.ac.in)

most common method used to obtain the cubic lithium garnets with high conductivity. Lanthanides doping at La sites have been studied but it does not have any significant effect on the conductivity of garnets[12]. Ce doping on LLZO at La sites has been attempted, however, a limited incorporation of Ce was detected[13]. Doping with Ce on the site of Zr in LLZO is an effective method due to the large size of the Ce^{4+} ion in comparison to Zr^{4+} ion, as doping reduces the tetragonal distortion of LLZO resulting in high ionic conductivity of material[14].

Moreover, the high ionic conductivity can also be obtained by various sintering techniques like spark plasma sintering and hot press sintering in controlled sintering atmosphere *etc*[14]. Despite of optimizing these parameters, reaction mechanism during the sintering period is important to understand. It has been found that during sintering some impurity phase like $La_2Zr_2O_7$, $LiAlO_4$ creates hindrance in ionic movement, that can be eliminated at 1130 ºC for 12 hrs. Hence, the investigation reveals that sintering is also an important factor to consider[15, 16]. To minimize the impurity phase and increase the ionic conductivity of LLZO, Xu *et al.*[17] designed multi-step holding strategy to produce high density garnets.

Therefore, the present study demonstrates a simple and efficient method used to produce compact ceramics with high Li^+ ion conductivity using a multi-step sintering process to synthesize lithium garnet with molecular formula: $Li_7La_3Zr_{1.75}Ce_{0.25}O_{12}$ (LLZCO) as solid state electrolyte material. The structural, microstructural analysis, ionic conductivities, electronic conductivities and electrochemical voltage stability of the LLZCO are also studied and improved to support the electrochemical usages further.

2 Material and Methods

2.1 Material synthesis

Synthesis of LLZCO has been attempted via solid-state reaction route. All precursors Li_2CO_3, CeO_2, ZrO_2, La_2O_3 with purity >99% were dried in Metrex instrument make oven at 120 ºC overnight to remove any moisture present. In order to account for the lithium loss at high temperatures, 10% more Li_2CO_3 was added. Fig. 1 depicts a schematic of the synthesis of LLZCO. All the precursors were ball-milled in zirconia jar using ethanol as milling media to produce the raw powder, which was then calcined at 900 ºC in a box furnace for 10 hrs. The calcined powder was uniaxially pressed into pellets of diameter, 10 mm and thickness, 1.0 mm. These pellets were then moved to furnace for multi-step sintering process in which initially the pellet was sintered at 1150 ºC for 2 hrs than at 1100 ºC for 2, 4, 6, 8 hrs marked as sample T1, T2, T3, T4, respectively, in this study.

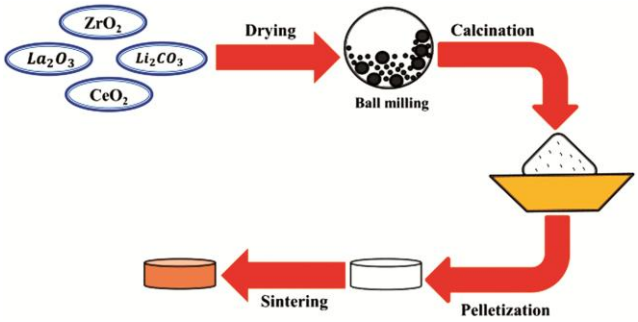


Fig. 1 — Schematic presentation of synthesis of LLZCO via solid state reaction route in ambient atmosphere.

2.2 Material characterization

As-synthesized sample was characterized by Thermogravimetric analysis (TGA) and Differential Scanning Calorimetry (DSC) to study the thermal stability and mass loss of the material via Perkin Elmer make model TGA 4000 equipment. Synthesized sample was characterized via X-ray diffractometer (Bruker make, model: D-8) using CuKα(λ=1.54Å) source in the range of 10-55° with a step size of 0.02°. Density of sintered pellets were calculated using simple physical process mass divided by volume. To study the microstructure of the sintered pellets using scanning electron microscopy (SEM), FEI make Nova Nano system has been used. The impedance spectroscopy analysis of LLZCO pellet has been carried out in the frequency range of 1 MHz to 4 Hz within the temperature range 30 and 70 °C using HIOKI make 3532 LCR meter. Hence, the estimation of ionic conductivity and activation energy were performed. Chronoamperometry test of dense pellet was done by biologic VMP3 potentiostat at a constant voltage of 100mV for 2 hrs to calculate its steady state electronic conductivity. Sintered pellet was coated by silver paste on both faces to make them electrically conductive. Electrochemical voltage stability was performed in the range of -0.5 V to 4.0 V using SS|LLZCO|Li cell, where stainless steel(SS) is used as working electrode and Li-metal as reference electrode. The cell fabrication was performed in MBraun make Ar filled Glove box workstation with O_2 and H_2O level less than 0.5 ppm.

3 Results and Discussion

3.1 Thermogravemetric(TGA) and differential scanning calorimetry(DSC) analysis

To study the phase formation and sintering temperature of LLZCO ball-milled product of the precursors was examined via TGA and DSC as shown in Fig. 2. TGA shows multi-stage mass loss occurs, initially 0.84% mass loss occurs that depicts the removal of moisture and reaction with La_2O_3 to form $La(OH)_3$. Further, mass loss of 2.01% and 2.25% accompanying with endothermic peaks (in DSC plot) at 335 and 450 °C can be attributed to La_2O_3 decomposition into LaOOH and then LaOOH reaction with CO_2, respectively[18]. Final mass loss of 17.44% starts from 700 °C, which indicate that the decomposition of Li_2CO_3 has been started leading to the formation of LLZCO[19]. It can be seen that above 1050 °C mass loss is negligible. Therefore, 1150 and 1100 °C has been considered for sintering temperature of material.

3.2 Structural analysis

X-ray diffraction analysis of an LLZCO sample sintered at 1150 °C for 2hrs is shown in Fig. 3 Stoichiometric amount of Ce doped $Li_7La_3Zr_2O_{12}$ can be indexed with cubic structure of space group: Ia-3d and it matched with standard JCPDS file: 01-080-6067. While extra peaks indicate the $La_{24}Ce_{12}Li_{24}O_{72}$ impurity phase[12]. The average crystallite size of LLZCO is calculated using Scherrer's formula which is given by Eq. 1. Fig. 4 shows the Williamson-Hall plot which is used to determine the micro-strain from the total peak broadening of diffraction peaks, using the expression stated in Eq. 2.

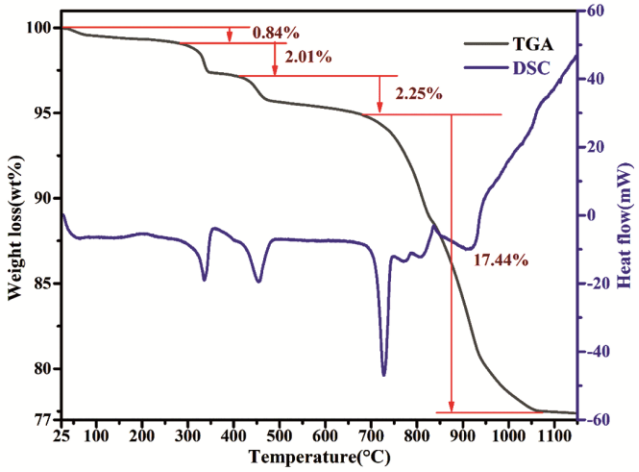


Fig. 2 — TGA/DSC curve of ball-milled product upto 1150 °C.

$$D = \frac{K\lambda}{\beta Cos\theta} \quad \ldots (1)$$

$$\beta_t = \frac{K\lambda}{DCos\theta} + 4\varepsilon tan\theta \quad \ldots (2)$$

Here, symbols β_t, β, λ, ε, D, and K are the total peak broadening, full width at half maxima, wavelength of X-ray source (1.54Å), micro-strain, crystallite size, and shape factor (0.9), respectively. High intense peaks in XRD are assigned to planes (211), (321), (400), (420), (422), (521), (611), (642) of LLZCO. The peaks are used to calculate the average full width at half maxima for crystallite size estimation. The micro-strain caused in the system is indicated by the slope of the fitted plot between 4sinθ and βcosθ. The average crystallite size calculated from Scherrer's formula is 100nm and micro-strain in crystal system obtained from the W-H plots is 3.55×10^{-4}, respectively. Smaller crystallite size is one of the reason for better sintering of a ceramic[20]. Here

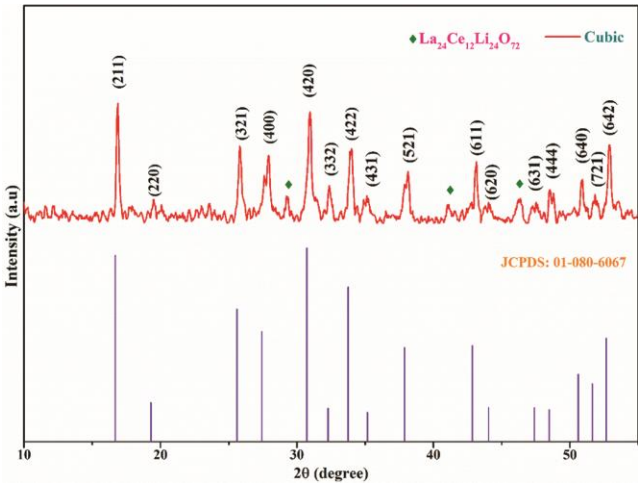


Fig. 3 — XRD pattern of LLZCO synthesized at 1150 °C for 2hrs.

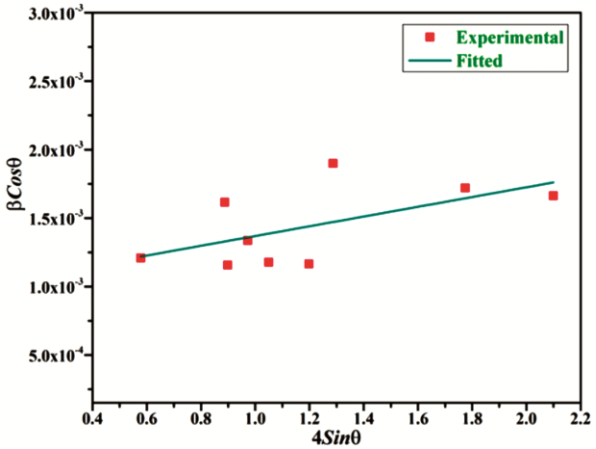


Fig. 4 — W-H plot of LLZCO sample.

in this study, it shows the sub-micron crystallite size range which is one of the reason for better sintering and lowering the sintering temperature of bare material[9].

3.3 Microstructure and density analysis

High density of the ceramic solid electrolyte is one of the major factor for the high conductivity and pure phase of the material. Therefore, LLZO pellet densification may prevent the growth of impurities. It is observed that at high temperature Al diffusion from alumina crucible can be diminished by higher relative density of the material which diminishes the impurity phase[16].

Fig. 5(a-d) display the SEM micrographs of LLZCO samplesT1, T2, T3 and T4 sintered initially at high temperature of 1150 °C for 2 hrs and then at 1100 °C for 2, 4, 6 and 8 hrs, respectively. The relative density results of these samples of material, LLZCO sintered at different schedules of temperatures and time are shown in Table 1. The difference in relative density of the material can be attributed to the grain growth and change in porosity of the material. Generally, in single step sintering from 900 °C to 1200 °C, grain size increases and porosity decreases which results in increment of relative density[16]. Among all the sintered samples, it can be analyzed that sample T2 reveals the highest relative density of 65%. The SEM images of different samples is shown in Fig. 5(a-d) which clearly shows that as the sintering time increases the grain boundary growth first increases and then decreases. Therefore, the increases/decrease in the density can be explained on the basis of movement of intergranular melt inside/outside of grains[17]. Fig. 5(b) shows that for sintering time of 4 hrs LLZCO showed interconnected and uniform grain distribution, which is in agreement with high relative density. The porosity increases as the exposure time increases more than 4 hrs as shown in Fig. 5(c-d). The higher sintering time may lead to lithium loss and reducing the grain size as uniform grain of sample T2 breaks into smaller flakes like grain for T3 and T4 samples. Fig. 6(a) shows the pellet before sintering and Fig. 6(b-e) show the sintered pellets of sample T1, T2, T3, T4, respectively. A reddish color in sample T4 pellet can be attributed to the lithium loss due to more sintering time. Fig. 6(f) shows the multiple production and reproducibility of density on 4-6 pellets of each sample. It can be seen that maximum density is

Table 1 — Multi-step sintering and Relative density of LLZCO samples.

S.no	Sample ID	Sintering Treatment	Relative Density
1.	T1	1150 °C for 2 hrs and 1100 °C for 2 hrs	44
2.	T2	1150 °C for 2 hrs and 1100 °C for 4 hrs	65
3.	T3	1150 °C for 2 hrs and 1100 °C for 6 hrs	61
4.	T4	1150 °C for 2 hrs and 1100 °C for 8 hrs	57

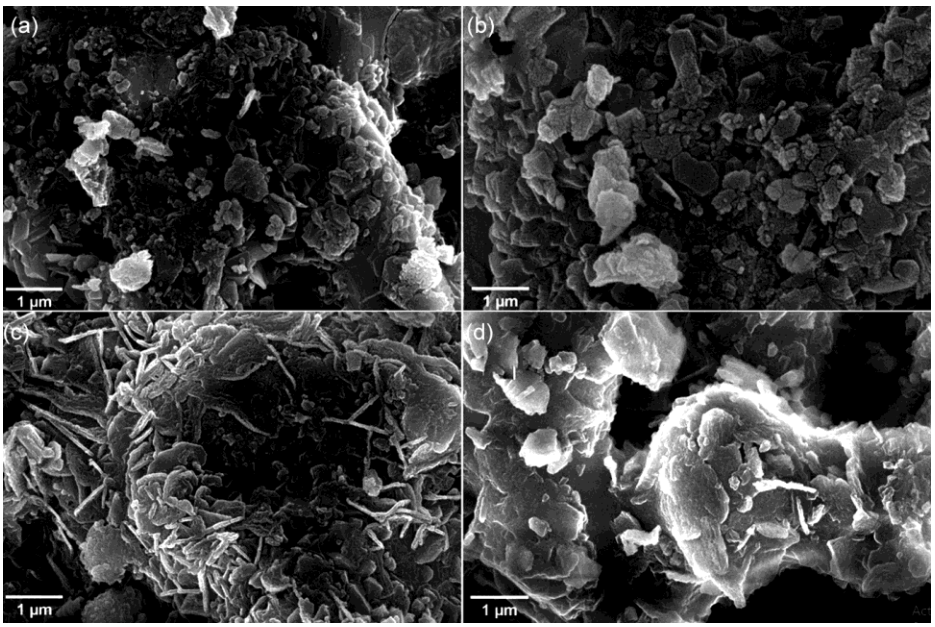


Fig. 5 — SEM micrograph of sample: (a)-T1, (b)-T2, (c)-T3 and (d)-T4 at a magnification of 25 KX.

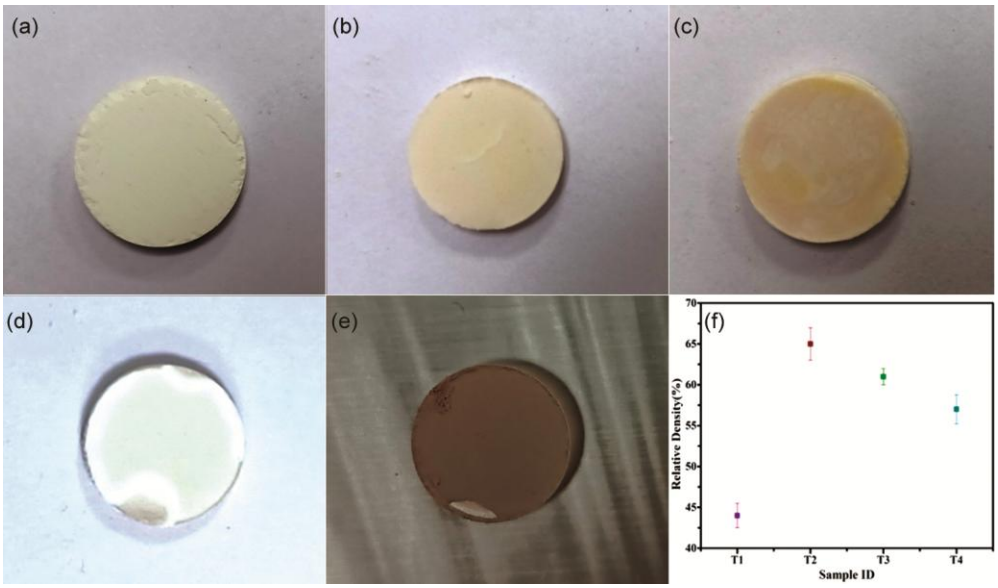


Fig. 6 — Before sintering (a); after sintering (b)T1, (c) T2, (d) T3, (e) T4, samples and (f)- variation of relative density in multiple attempts.

achieved for sample T2. SEM images of T2 sample is shown in Fig. 7, LLZCO pellets have the dense microstructure with average grain size of 408nm. Fig. 8(a-f) shows the elemental mapping of LLZCO which confirms almost homogeneous distribution of O, Zr, La, Ce, respectively, throughout the sample. The rise in intergranular melt with temperature is the cause of the increase in relative density, improved garnet phase and dense microstructure while multi-step sintering indicates the further densification of garnet at high temperature sintering[14].

3.4 Conductivity and activation energy analysis

Impedance spectroscopy measures sample's impedance over a broad frequency range and uses that information to characterize the material's various components based on their electrical relaxation times. Impedance is a combination of resistance, inductance and capacitance which can be calculated by passing AC signal through the sample[24]. Fig. 9 shows the impedance plot of Z' vs Z''pellet of LLZCO sample T2 sintered at 1150 °C for 2 hrs and then 1100 °C for 4hrs measured at 30 °C (room temperature). The semicircles observed in the high frequency region indicate the bulk and grain boundary contribution. Hence, each semicircle can be explained as equivalent to parallel combination of constant phase elements and resistance as shown in equivalent circuit of the inset of Fig. 9. Overall impedance is fitted via equivalent circuit of (R1)(R2Q1)(R3Q2)(Q3) using a EIS analyzer software, where R1, R2, R3 are electrode resistance, bulk resistance and grain boundary resistance respectively and Q1, Q2, Q3 are the constant phase elements. Hence, computed values for R1, R2, R3, Q1, Q2, Q3 are 2120 Ω, 7082 Ω,17050 Ω, 55.35×10^{-9} F, 2.11×10^{-9} F and 12.58×10^{-6} F, respectively, for impedance plot at room temperature. The ionic conductivity of the sintered pellets is calculated using the Eq. 3;

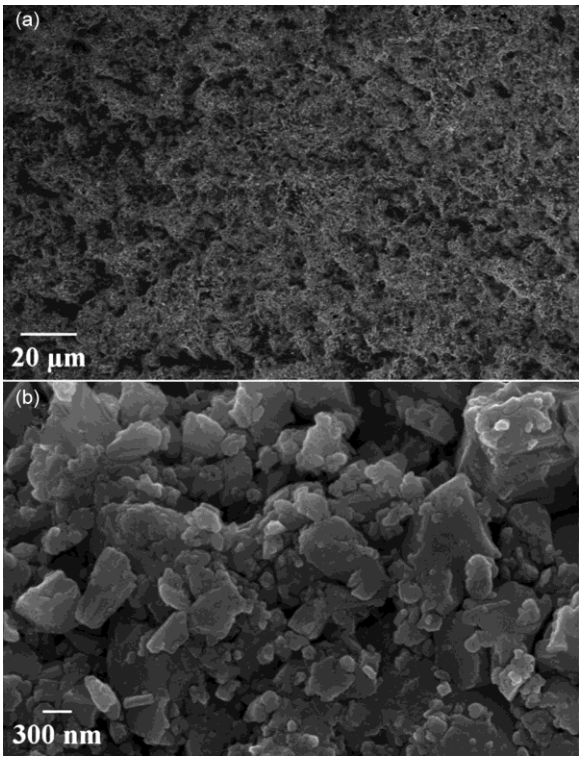


Fig. 7 — SEM micrograph of sample T2 pellet (a) 500 X (b) 45 KX.

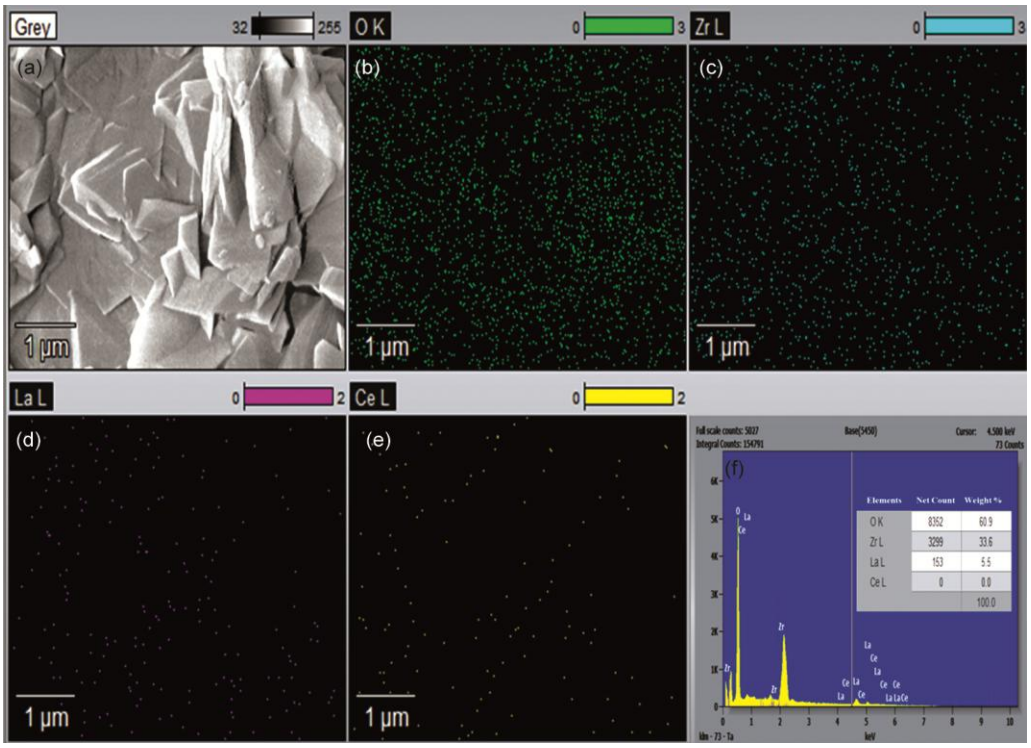


Fig. 8 — (a):SEM micrograph of sample T2; Elemental mapping images of (b)- O(green), (c)-Zr(blue), (d)- La(magenta), (e)- Ce(yellow) of LLZCO sample T2, (f) Result of EDS analysis of LLZCO.

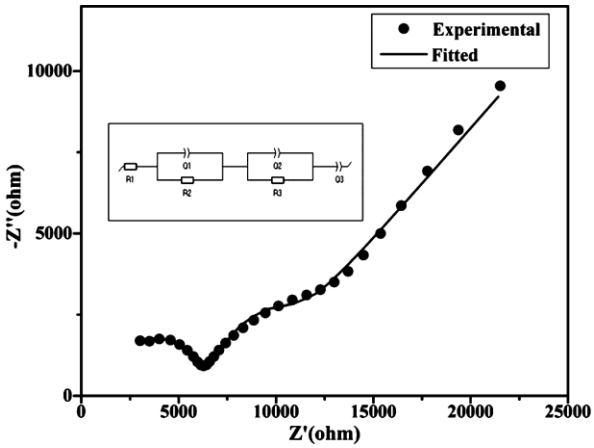


Fig. 9 — Z' Vs Z'' curveof LLZCO sample T2 pellet sintered at 1100 °C for 4 hrs at 30 °C.

$$\sigma = \frac{L}{R \times A} \quad \text{... (3)}$$

where R is the resistance of the sample in Ω, A is the area cm^2 and L is the thickness of the pellet in cm, respectively. The bulk ionic conductivity (σ) of the LLZCO material is observed as $0.17\times10^{-4}Scm^{-1}$ at 30 °C. Fig. 10 shows the impedance plots Z' vs Z'' at different higher temperatures ranging from 30 to 70 °C for sample T2. Here, it is observed that as the temperature increases both the contribution (bulk and grain boundary) merges. Therefore, the sum of the bulk and grain-boundary resistance contribution is considered for ionic conductivity calculation. Fig. 11 shows the temperature dependence of the conductivity estimated using Arrhenius Eq. 4 and given by;

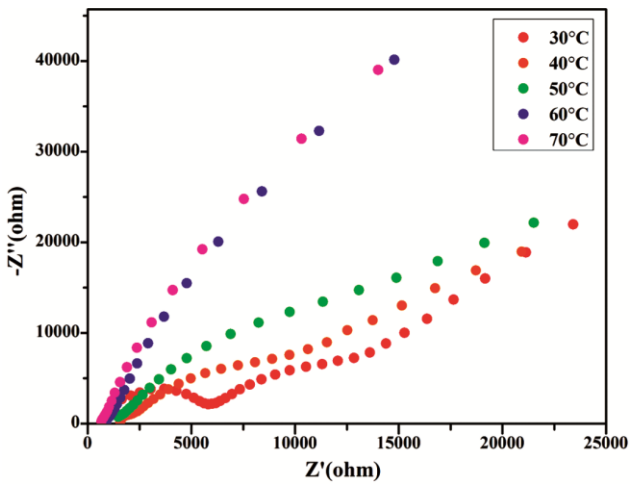


Fig. 10 — Temperature dependent impedance spectroscopy of LLZCO sample T2 pellet sintered at 1150 °C for 2 hrs and 1100 °C for 4 hrs from 30 °C to 70 °C.

$$\sigma = \sigma_0 e^{-E_a/K_B T} \quad \text{... (4)}$$

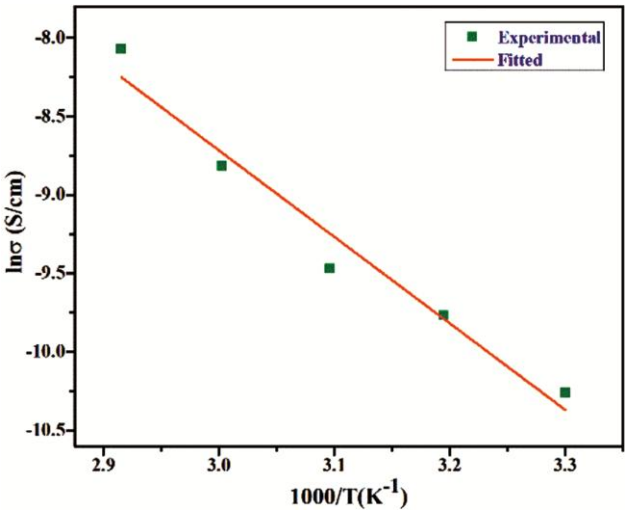


Fig. 11 — Activation energy curve of LLZCOsample T2 pellet sintered at 1150 ºC for 2 hrs and 1100 ºC for 4 hrs from 30 °C to 70 °C.

where σ, σ_o, E_a, T, K_B are ionic conductivity, pre-exponential factor, activation energy, temperature and Boltzmann constant, respectively. The activation energy was estimated to be 0.47 eV. Activation energy and ionic conductivity has a strong correlation between them as the activation energy decreases ionic conductivity increases while at the same time it has been seen that the Li^+diffusibility increases across the material due to lower in activation energy, high diffusibility of ions may enhance the cyclability of solid state battery[21]. Here, Ce stabilized LLZO activation energy(0.47 eV) value can be seen as of the decrease in activation energy from bare LLZO[10]. Fig. 12 shows the chronoamperometry curve of LLZCO pellet which is used to calculate the electronic conductivity of solid electrolyte using Eq. 5;

$$\sigma_e = \frac{I_{SS} \times L}{V \times A} \quad \ldots (5)$$

Where σ_e, I_{SS}, I_o, L, V, A are electronic conductivity, steady state current, thickness of pellet, polarization voltage, area of pellet, respectively[22]. It can be seen that the value of initial output current (I_o) of LLZCO sample is 2.33×10^{-5} mA, after two hours of constant voltage input sample shows very low value of steady state current (ISS), 5.24×10^{-6} mA, it leads to the negligible electronic conductivity (σ_e), 1.00×10^{-8}Scm^{-1} of LLZCO ceramic solid state electrolyte. Lower values of σ_e is very crucial for battery application as high electronic conductivity, σ_e favour the dendrite formation in Li^+ batteries which results in short circuiting of battery, decrease in coulombic efficiency

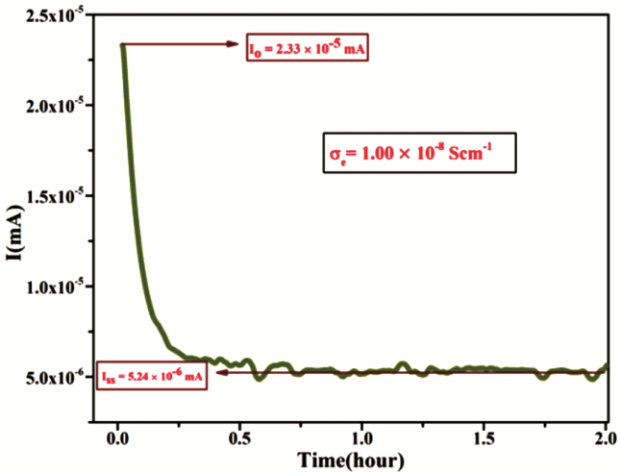


Fig. 12 — Chronoamperometry of LLZCO(T2) pellet sintered at 1150 °C for 2 hrs and 1100 °C for 4 hrs.

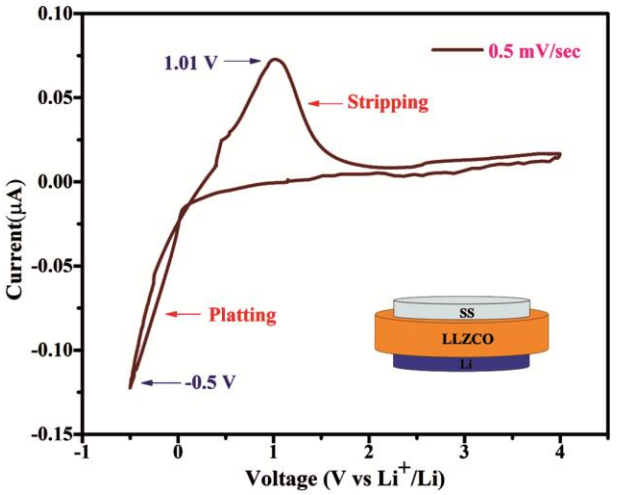


Fig. 13 — Cyclic Voltammetry analysis of LLZCO sample T2 pellet sintered at 1150 °C for 2 hrs and 1100 °C for 4 hrs solid electrolyte in the range of -0.5 V to 4 V.

and capacity loss[23]. Ce doping in LLZO is helpful to stabilize cubic phase at lower temperature as compared to undoped LLZO[9]. Hence, multi-step sintering approach leads to the formation of dense ceramic and yields high conductivity then undoped LLZO tetragonal phase. These alterations are beneficial for solid electrolyte application[10].

3.5 Electrochemical Analysis

The cyclic voltammetry (CV) analysis of LLZCO of sample T2 has been attempted to test the electrochemical stability using SS|LLZCO|Li cell configuration. Here, Stainless Steel (SS) is used as working electrode and Li metal chip is used as reference electrode. Fig. 13 shows the CV curve ranging from -0.5 to 4.0 V at a scan rate of 0.5 mV/s.

The peaks observed at -0.5 V and 1.01 V corresponds to Li platting and stripping, respectively. The broadening of peaks can be attributed to additional redox process taken place due to cerium, Ce_4O_7 phase formation and because of reaction at the interface or insertion of excessive Li in the electrolyte structure[14]. Furthermore, curve shows that there is no other extra oxidation peak upto 4.0 V which indicates that LLZCO has wide and stable electrochemical window to lithium metal and can be used for high energy density Li-ion batteries applications.

4 Conclusion

In this study, a garnet type $Li_7La_3Zr_2O_{12}$ is successfully synthesized via solid state reaction route. TGA/DSC analysis has been carried out upto 1150 °C to check the mass loss and optimum temperature range for multi-step sintering.TGA shows overall 22% mass loss has been occurred till 1050 °C and exothermic peak above 950 °C shows the crystallization has been started. Hence, 1150 and 1100 °C temperatures has been used for multi-step sintering. XRD analysis confirms that LLZCO has a garnet-type structure (space group: Ia-3d) with some impurities of the $La_{24}Ce_{12}Li_{24}O_{72}$. High temperature sintering at 1150 °C for 2 hrs followed by 1100 °C for 4 hrs shows the highest relative density, dense microstructure and uniform grain distribution of synthesized LLZCO solid state electrolyte. The bulk Li-ion conductivity (σ) of as synthesized LLZCO solid electrolyte is observed, $0.17\times10^{-4}Scm^{-1}$ along with low electronic conductivity (σ_e) of $1.00\times10^{-8}Scm^{-1}$. Minimum activation energy of the optimized sample T2 of LLZCO is observed as 0.47 eV. Electrochemical analysis by CV result shows platting peaks of lithium ion at -0.5 V and broad stripping peak for lithium ion is due to Ce_4O_7 phase formation. Hence, no other peak has been detected upto 4 V in CV plot, hence the electrolyte is stable upto 4.0 V.

Acknowledgment

Authors acknowledge Delhi Technological University, Delhi to provide financial support through a project grant “DTU/IRD/619/2019/2114” to carry this research work.

References

1 Larcher D & Tarascon J M, *Nature Chem*, 20 (2018) 19.
2 Mrinalini M, Hsu C, Rath P C, Patra J, Lai H Z, Chang T L, Wang C, Wu T, Lee T & Chang J, *Electrochimica Acta*, 353 (2020) 136536.
3 Lebedeva N P & Boon L, *J Electrochem Soc*, 163 (2016) A821.
4 Wu B, Wang S, Lochala J, Desrochers D, Liu B, Zhang W, Yang J & Xiao J, *Energy Environ Sci*, 11 (2018) 1803.
5 Liu B, Zhang J & Xu W, *Joule*, 2 (2018) 833.
6 Samson A J, Hofstetter K, Bag S & Thangadurai V, *Energy Environ Sci*, 12 (2019) 2957.
7 Meesala Y, Jena A, Chang H & Liu R, *ACS Energy Lett*, 2 (2017) 2734.
8 Thangadurai V, Kaack H & Weppner W J F, *J Am Ceram Soc*, 86 (2003) 437.
9 Murugan R,Thangadurai V & Weppner W, *Angewandte Chem Int Edn*, 46 (2007) 7778.
10 Awaka J, Kijima N, Hayakawa H & Akimoto J, *J Solid State Chem*, 182 (2009) 2046.
11 Jiang C, Li H & Wang C, *Sci Bull*, 62 (2017) 1473.
12 Abin K, Seungjun W, Minseok K, Heetaek P & Byoungwoo K, *Front Chem*, 8 (2008).
13 R Ezhiylmurugan, W Jeff, A Jan & S Jeff, *J Power Source*, 230 (2013) 261.
14 Dong B, Yeandel S R, Goddard P & Slater P R, *Chem Mater*, 32 (2019) 215.
15 Henrik B, Janis D, Stefan B, Alexander K, Patrick B, Martin W, Paul H, Anatoliy S, Helmut E, Andriy L, Viola D, Lorenz K & Jürgen J, *Phys Chem Chem Phys*, 13 (2011) 19378.
16 Xia W, Xu B, Duan H, Guo Y, Kang H, Li Hua & Liu Hezhou, *ACS Appl Mater Interfaces*, 8 (2016) 5335.
17 Xu B, Duan H, Xia W, Guo Y, Kang H, Li H & Liu H, *J Power Source*, 302 (2016) 291.
18 Janez K, Seyedabolfazl M, Benjamin W, Eeva-Leena R & Tanja K, *Solid State Ionics* 380 (2022) 115943.
19 Marlou K, Battista A G, Guk-Tae K, Varvara S, Meike S, Jörg S, Andreas R & Stefanoi P, *J Power Source*, 353 (2017) 287.
20 Andrew R & Ian C, *Metal-Reinforced Ceram*, (2021) 211.
21 Manju S & Yashonath S, *The J Chem Phys*, 129 (2008) 144103.
22 Ramasubramonian D & Reza S, *Batteries Supercaps*, 4 (2021) 596.
23 Fudong H, Andrew W S, Jie F, Fei W, Miaofang C, Donovan L, Nancy D, Howard W & Chunsheng W, *Nature Energy*, 4 (2019) 187.
24 Brugge, Rowena, *Garnet ceramic electrolytes for next-generation lithium batteries*, University of London, 2018.

AIP Conference Proceedings

RESEARCH ARTICLE | FEBRUARY 20 2024

Effect of rare earth metal (lanthanum) doping on optical and electronic properties of lithium niobate ($LiNbO_3$): DFT insights FREE

Ashish Raturi ✉; Poornima Mittal; Sudhanshu Choudhary

Check for updates

AIP Conf. Proc. 2986, 030060 (2024)
https://doi.org/10.1063/5.0192542

View Online
Export Citation
CrossMark

26 February 2024 07:49:42



AIP Publishing

Effect of Rare Earth Metal (Lanthanum) Doping on Optical and Electronic Properties of Lithium Niobate ($LiNbO_3$): DFT Insights

Ashish Raturi[1, a)], Poornima Mittal[1, b)] and Sudhanshu Choudhary[2, c)]

[1]*Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India*
[2]*Department of Electronics and Communication Engineering, National Institute of Technology, Kurukshetra, India*

[a)] Corresponding author: ashish.raturi121@gmail.com
[b)] poornimamittal@dtu.ac.in
[c)] sudhanshu@nitkkr.ac.in

Abstract. In this paper the optical properties and electronic strcuture of rare earth metal lanthanum doped lithium niobite are presented. All the computations are achieved on ATK VNL tool kit by utilizing density functional theory. In electronic properties the computations are performed to calculate the band gap of the lanthanum doped lithium niobite is calculated. For optical properties, dielectric function, refractive index and reflection coefficient are calculated. The undoped lithium niobite is a wide band material with absorption in UV region. It is illustrated from the results obtained that value the band gap is decreased significantly due to doping with lanthanum, which shifts the absorption in the higher wavelengths region in compression of the undoped crystal of lithium niobite. The shift in the absorption towards higher wavelengths is identified as red shift. The decrease in the band gap and relocation of the absorption in the higher wavelengths opens up an opportunity of utilizing the lanthanum doped lithium niobite for the optoelectronic applications.

Keywords: optoelectronic, DFT, perovskites, absorption

26 February 2024 07:49:42

INTRODUCTION

In the wide availability of the materials, materials with ferroelectric properties have proved them promising for the optoelectronic applications [1-3]. Theses material are fascinating for the various applications due to their extraordinary and versatile piezoelectric and optical properties. Lithium niobite widely known as LN, is a famous and widely acceptable material for the optoelectronic applications, because of its excellent optoelectronic properties [4-5]. $LiNbO_3$ is a perovskite with the formula ABO3, in which A is the lithium atom an B is niobite atom. LN is usually found in hexagonal crystal structure and researched by the several scholars due to its wide optical, piezoelectric, photonic and electro-optic properties. It has wide acceptability in the manufacturing of optical storage, modulators, wave guides and acoustic devices [6].

The $LiNbO_3$ is a wide band gap material with a band gap of value 3.56 eV [7]. Due to wide band gap the non-zero values of the absorption for the undoped $LiNbO_3$ is available in the ultra-violate (UV) region only. As the UV region is only 5% of the sun light, a significantly high proportion of the sun light remains unutilized [8]. This is not favourable for the utilization of the materials for photonic and optoelectronic applications. In order to shift the absorption in the visible wavelength, several techniques like doping, strain, heterostructure are utilized by the researchers past to tune the optoelectronic properties of various materials and structures [9-10]. Among all, doing proved itself and accepted by the researchers as an efficient way to optimize the optoelectronic properties of the various structures and materials [11-13]. In past, researchers calculated and examined the various optical properties of pure and doped $LiNbO_3$ for various photonic and optoelectronics applications. Hossain et al. calculated the optoelectronic properties of $LiNbO_3$ and reported the band gap and other optical parameters of $LiNbO_3$ using DFT

3rd International Conference on Functional Materials, Manufacturing, and Performances
AIP Conf. Proc. 2986, 030060-1–030060-5; https://doi.org/10.1063/5.0192542
Published by AIP Publishing. 978-0-7354-4840-7/$30.00

[7]. M Arshad Javid et al. reported various properties of $LiNbO_3$ using GGA-PBE [2]. Further, A Raturi et al. examined the effect of doping (metal dopant) on the optical response of $LiNbO_3$ [5]. X Liu et al. studied the consequence of CU doing on the optical parameters of $LiNbO_3$ [14]. In the studies, it has been determined that the doping is an efficient and widely acceptable way of improving the optoelectronic properties of $LiNbO_3$ for various applications.

In this paper to examine the consequence of doping of lanthanum on optoelectronic properties of $LiNbO_3$. The undoped hexagonal cell of $LiNbO_3$ is doped with rare earth metal lanthanum and further the electronic stricture and optical parameters like band gap, dielectric constant, refractive index, and absorption coefficient are computed to comprehend the optical behaviour of the material. Values of all the optical parameters are computed for different wavelength and energies.

COMPUTATIONAL METHODOLOGY

To calculate the optoelectronic properties of lanthanum doped $LiNbO_3$, a hexagonal cell of $LiNbO_3$ is doped with lanthanum as shown in figure 1. All the properties are calculated using density functional theory calculations performed on ATK VNL. The geometry of doped $LiNbO_3$ is relaxed using PBE of GGA [15]. The geometry is relaxed till the forces between the atoms reaches 0.02 eV/A and the stress is reached to a value 0.002 GPa [15]. The sample point of 5x5x1 are taken for all the calculations with Monkhorst-Pack sampling. The adopted value of the mesh cut off is 150 Ry. The bandgap of LN is computed using GGA. However, m-GGA functional is taken up for the calculations of optical properties [16-18].

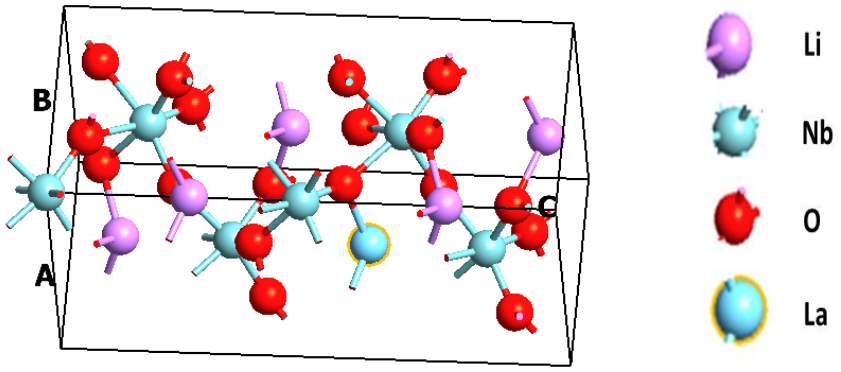


FIGURE 1. Hexagonal cell of Lanthanum doped $LiNbO_3$.

RESULTS AND ANALYSIS

In this segment of the paper, the structural and electronic properties of lanthanum (La) doped $LiNbO_3$ are calculated and examined to underrated the effect of La doping.

Structural and Electronic Properties

For examining the properties of lanthanum doped $LiNbO_3$, a crystal of undoped $LiNbO_3$ is doped with lanthanum with doping concentration of ~10%. The computed values of the lattice constants after optimization of structure of the undoped $LiNbO_3$ cell are a=b= 5.185 Å, c=13.8Å [5]. However, after the doping the lattice constants become a=b= 5.165 Å, c=13.4 Å. The bandgap of the la-doped $LiNbO_3$ is computed to comprehend the effect of doping on electronic configuration. Band gape effectively decide the characteristics of the materials and governs the properties like carrier concentration and absorption. The materials with low value of band gap can absorb the light of high wavelength. The materials with wide band gap absorb the light with low wavelength (UV) and require high energy to jump from valance band to conduction band. $LiNbO_3$ is a wide band gap material with the band gape of 3.5 eV. For $LiNbO_3$, the absorption is available barely in the UV wavelengths [7]. After doping the band gap of the

26 February 2024 07:49:42

La-doped $LiNbO_3$ is calculated and is shown in figure 2. The band gap of the La doped $LiNbO_3$ is 2.71 eV, which is lowered down in comparison of the undoped crystal because of doping. The lessening of the bandgap is as a consequence of the change in chemical composition. The dopants create energy states in the available forbidden gap and generates a band or multiple bands. These bands are close to the conduction and valance band, which ultimately results in decrease of band gap.

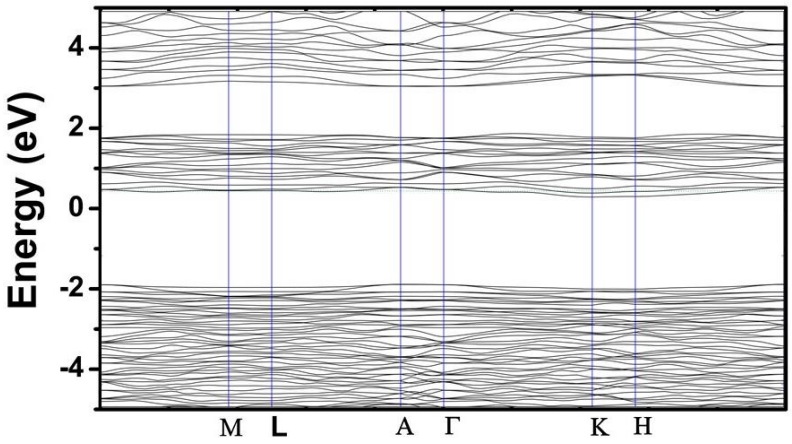


FIGURE 2. Band structure of Hexagonal cell of Lanthanum doped $LiNbO_3$.

Optical Properties

In this segment of the paper, various optical properties of the La-doped $LiNbO_3$ are calculated and examined to explore the effect of doping with La on the optical behavior. In optical properties we calculate absorption coefficient as a function of energy and wavelength. Which is the representation of the amount of light the material will absorb for different wavelength. Further, the dielectric constant as a function of energy is calculated to comprehend the losses. The dielectric constant (real portion) signifies the slowdown of the traveling of light in the medium and imaginary part is the demonstration of damping.

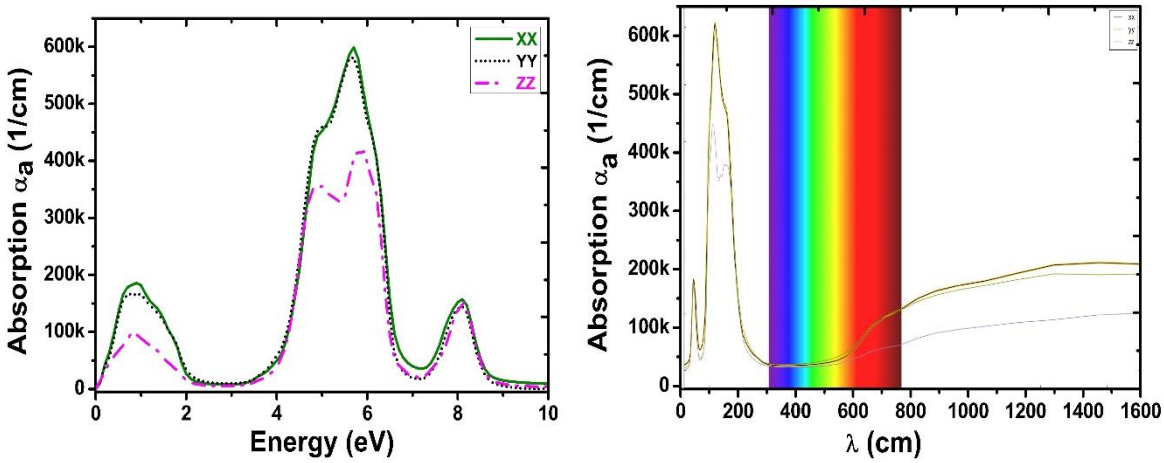


FIGURE 3. Absorption of Lanthanum doped $LiNbO_3$.

26 February 2024 07:49:42

The absorption coefficients of the La doped $LiNbO_3$ as a function of energy and wavelength are illustrated in figure 3. For undoped $LiNbO_3$, the absorption is restricted to the UV region. The values of absorption are non-zero for the lower energies with no values of the absorption in the visible wavelength area[7]. For La-doped $LiNbO_3$, as illustrated in figure 3, the absorption is increased for the visible wavelengths, showing red shift. For La doped $LiNbO_3$,the optical absorption has non-zero values from ~3 eV, which is equal to the calculated band gap. The absorption is in the UV range along with increased values of the absorption in the rage 600 to 800 nm (visible region). The phenomena of red shift in La-doped $LiNbO_3$ is the consequence of the exchange interaction in the charge carriers and due to the tapering of band gap, which cause a built-in electric field.

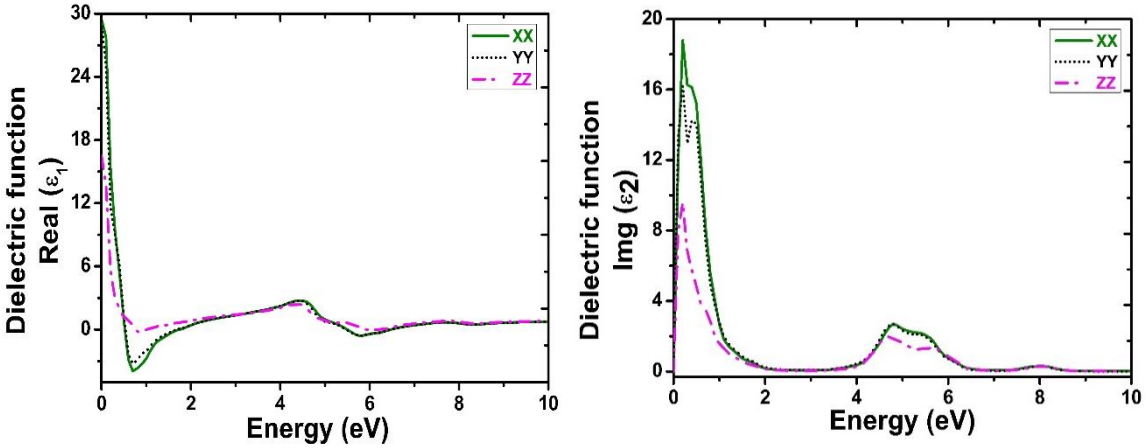


FIGURE 4. Dielectric function Lanthanum doped $LiNbO_3$.

The energy dependent dielectric function for La doped $LiNbO_3$ is illustrated in figure 4. The dielectric constant (imaginary) is the demonstration of the consequence of the damping and is well-considered as an efficient method of determining the absorption. The value of dielectric constant (imaginary part) is increased due to doing, which results in better optical absorption. The real dietetic constant explains the phase lag and the value is increased due to doping, which results in better spread of the light through $LiNbO_3$.

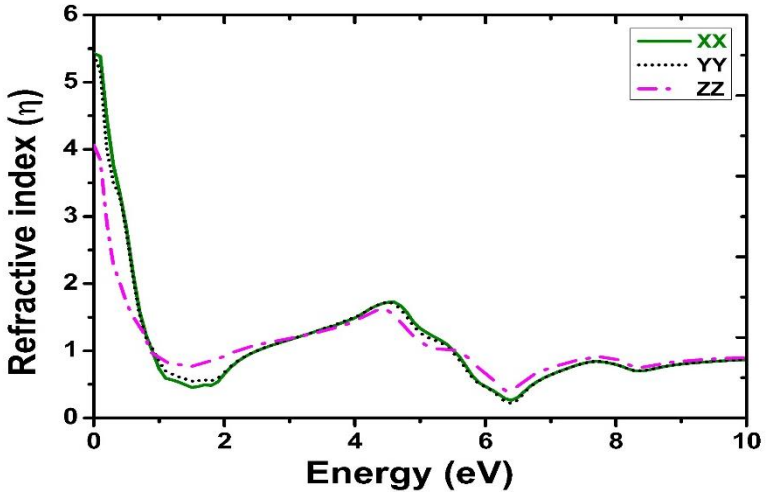


FIGURE 5. Refractive index of Lanthanum doped $LiNbO_3$.

26 February 2024 07:49:42

The refractive index (function of energy) for la doped $LiNbO_3$ is illustrated in figure 5. The values of energy dependent refractive index are calculated from the dielectric constant (real part). The value of intrinsic refractive constant is increased because of the La-doping, which ultimately results in the extended life time of the light traveling through $LiNbO_3$.

CONCLUSION

The optoelectronic properties of rare earth metal (lanthanum) doped lithium niobite crystal are computed and investigated. The undoped hexagonal structure of lithium niobite crystal is doped by the rare earth metal lanthanum and various properties are computed by utilizing density functional theory. For doped lithium niobite the value of band gap is 2.71 eV, which is low with respect to the undoped LiNbO3 crystal. The narrowing of the band gap is desired for the various optoelectronic applications. For the doped lithium niobite the absorption is remarkably shifted to the region of the lower energy region or higher side of wavelengths. This phenomenon is widely known as red shift and is favourable for the optoelectronics applications. Further other optical properties like dielectric constant, refractive index is also calculated to comprehend the optical behaviour of the lanthanum doped lithium niobite. The results strongly indicate that the lanthanum doped lithium niobite may be used as a possible candidate for the optoelectronic applications.

REFERENCES

1. R. Husin, F.W. Badrudin, M.F.M. Taib, and M.Z.A. Yahya, Mater. Res. Express **6**, 114002 (2019).
2. M. Arshad Javid, Z.U. Khan, Z. Mehmood, A. Nabi, F. Hussain, M. Imran, M. Nadeem, and N. Anjum, Int. J. Mod. Phys. B 32, 1850168 (2018).
3. X. Liu, W. Que, Y. He, and H. Zhou, J. Adv. Dielectr. **8,** 1820002 (2018).
4. A. Raturi and P. Mittal, in 2020 Int. Conf. Electr. Electron. Eng. (2020), pp. 597–600.
5. A. Raturi, P. Mittal, and S. Choudhary, Bull. Mater. Sci. **44**, 1 (2021).
6. A. Navrotsky, Chem. Mater. **10**, 2787 (1998).
7. M.M. Hossain, Heliyon 5, e01436 (2019).
8. B. Kumar, B.K. Kaushik, Y.S. Negi, and P. Mittal, in 2011 Int. Conf. Emerg. Trends Electr. Comput. Technol. (2011), pp. 702–707.
9. Y. Bi, M.F. Ehsan, Y. Huang, J. Jin, and T. He, J. CO2 Util. 12, **43** (2015).
10. J. Lim, H. Lim, and Y.S. Lee, Curr. Appl. Phys. **19**, 1177 (2019).
11. Negi, P. Mittal, and B. Kumar, Microsyst. Technol. **24,** 4981 (2018).
12. P. Mittal, J. Soc. Inf. Disp. **29**, 642 (2021).
13. T. Yamamoto, S. Suzuki, K. Kawaguchi, and K. Takahashi, Jpn. J. Appl. Phys. 37, 4737 (1998).
14. P. Mittal, S. Yadav, and S. Negi, Mater. Sci. Semicond. Process. 133, 105975 (2021).
15. R. Kochar and S. Choudhary, IEEE J. Quantum Electron. **54,** 1 (2018).
16. M. Tyagi and S. Choudhary, IET Optoelectron. **15**, 270 (2021).
17. S. Negi, P. Mittal, and B. Kumar, Recent Patents Eng. **14**, 373 (2020).
18. P. Mittal, Y.S. Negi, and R.K. Singh, in Adv. Mater. Res. (2013), pp. 585–589.

26 February 2024 07:49:42

Computers and Electrical Engineering 115 (2024) 109130

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Efficient prediction of coronary artery disease using machine learning algorithms with feature selection techniques☆

Md. Mehedi Hassan [a,1], Sadika Zaman [b,1], Md. Mushfiqur Rahman [c], Anupam Kumar Bairagi [a], Walid El-Shafai [d,e], Rajkumar Singh Rathore [f], Deepak Gupta [g,*]

[a] Computer Science and Engineering Discipline, Khulna University, Khulna, 9208, Bangladesh
[b] Department of Computer Science and Engineering, North Western University, Khulna, 9100, Bangladesh
[c] Department of Statistics, University of Dhaka, Dhaka, 1000, Bangladesh
[d] Security Engineering Lab, Computer Science Department, Prince Sultan University, Riyadh, 11586, Saudi Arabia
[e] Department of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf, 32952, Egypt
[f] Department of Computer Science, Cardiff School of Technologies, Cardiff Metropolitan University, Llandaff Campus, Western Avenue, Cardiff CF5 2YB, United Kingdom
[g] Department of Computer Science, Maharaja Agrasen Institute of Technology, India

ARTICLE INFO

Keywords:
Coronary artery disease
Cardiovascular health
Machine learning algorithms
Risk assessment
Feature selection
Medical data analysis
XGBoost
SHARP analysis

ABSTRACT

In recent years, there has been a notable surge in the prevalence of cardiovascular diseases (CVD), presenting a significant global public health challenge and a leading cause of mortality worldwide. Among the myriad complications stemming from CVD, heart failure stands out as a critical concern. Addressing heart failure through surgical means poses considerable challenges. The primary objective of this research is to identify pivotal attributes linked to heart failure and employ diverse machine learning methodologies to predict its occurrence, thereby enabling early estimation of mortality rates associated with heart failure. Leveraging a heart failure dataset, we conducted comprehensive model construction using pre-processing techniques such as feature scaling and correlation analysis. The Extreme Gradient Boosting (XGBoost) method was instrumental in evaluating feature relevance, leading to the selection of two distinct datasets: the whole dataset and the XGBoost dataset. In conclusion, we employed thirteen machine learning methods to predict the occurrence of death events within these datasets. Fine-tuning hyperparameters significantly enhanced model performance. Notably, our model demonstrated exceptional performance on this dataset, achieving the highest accuracy of 85.23% with Random Forest on the whole dataset and 86.36% with Flexible Discriminant Analysis on the XGBoost dataset.

☆ This paper is for regular issues of CAEE. Reviews were processed by Guest Editor Dr. Gaurav Garg and recommended for publication.

* Corresponding author.
E-mail addresses: mehedihassan@ieee.org (M.M. Hassan), sadikazaman@yahoo.com (S. Zaman), mushfiq.dustat@gmail.com (M.M. Rahman), anupam@ku.ac.bd (A.K. Bairagi), walid.elshafai@el-eng.menofia.edu.eg (W. El-Shafai), rsrathore@cardiffmet.ac.uk (R.S. Rathore), deepakgupta@mait.ac.in (D. Gupta).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.compeleceng.2024.109130
Received 24 July 2023; Received in revised form 12 February 2024; Accepted 13 February 2024
Available online 15 February 2024
0045-7906/© 2024 Elsevier Ltd. All rights reserved.

1. Introduction

Heart failure (HF) is a prominent contributor to mortality on a global scale. The phenomenon is becoming more prevalent daily. This disorder emerges when the cardiac system's capacity to distribute blood to all regions of the body is impaired. Cardiovascular disease is the primary cause of mortality on a global scale, resulting in the deaths of around 17.9 million people annually. CVD includes a range of medical conditions, such as coronary artery disease, cerebrovascular disease, rheumatic heart disease, and several other disorders affecting the heart and blood vessels. CVD often leads to fatalities mostly attributed to myocardial infarctions and heart failure, with a notable proportion of 33% transpiring among those below the age of 70 years [1]. The European Cardiology Society (ESC) reports that the worldwide prevalence of heart disease is estimated to impact 26 million people, and there is an annual incidence of 3.6 million new cases. Cardiovascular illness has been shown to result in a mortality rate of 50% among patients during the first two years after diagnosis. Additionally, about 3% of the total expenditure on healthcare is allocated towards the treatment of this condition. Males have a higher prevalence of heart disease compared to females, particularly in advanced age, but it may also manifest in children [2]. Multiple research studies have shown smoking, aging, high or low blood pressure, obesity, diabetes, and physical inactivity as significant risk factors for many illnesses. Analysts assert the significance of distinguishing between the many disorders linked to cardiovascular disease. The presence of venous obstruction might potentially indicate the presence of coronary artery disease, hence increasing the likelihood of experiencing heart failure, angina, or a cerebrovascular accident. There are several characteristics that may serve as indicators of the presence of CVD [3].

A study was undertaken with the objective of determining the most optimal feature selection method for the prognosis of cardiovascular disease. The XGBoost method demonstrated a classification accuracy of 73.74%, while the Support Vector Classifier (SVC) and Artificial Neural Network (ANN) attained accuracy rates of 73.18% and 73.20% respectively [4]. Another paper used a feature selection method grounded in rough sets in combination with the RF algorithm to provide an accuracy rate of 84% in their investigation [5]. In the research, a mix of Naive Bayes (NB), Random Forest (RF), Multilayer Perceptrons (MLP), and Bayesian networks was used, utilizing the majority voting (MV) technique. A precision level of 85.48% was achieved [6]. Numerous researchers are now engaged in the endeavor of using machine learning models to forecast the occurrence of cardiac disease. However, it is crucial to present an efficient and reliable model for tasks of this kind. Our study is focused on the optimization of characteristics and the classification of individuals with cardiac conditions. The primary contributions of this study are outlined as follows:

- We have compiled a dataset of significant features from heart patients for this study.
- Feature relevance has been determined utilizing XGBoost in order to prevent overfitting and underfitting.
- In order to assess the efficacy of the model using the datasets, two distinct datasets are generated: the complete dataset and the XGBoost dataset.
- Diverse machine learning methodologies are employed to analyze the training dataset and make predictions regarding heart failure.

Section 3 of our paper outlines the system's design technique. Dataset description, machine learning techniques employed in this research, and a description of the characteristics are all included in Section 4.'s implementation of our suggested work. The findings and conclusions of our research have been discussed in detail in Section 5. Sections 6 and 7 discuss the comparison table of our suggested model's accuracy with the current model and the comparative development of our proposed models. The study's findings and future research and limitations are discussed in Section 8.

2. Literature review

Researchers used both to predict the patient's survival and to rank the features that are most predictive of illness, according to Davide et al. [7]. Feature-ranking research using biostatistics approaches and traditional machine learning has been conducted, and the results of these two methods have been compared. They have also employed gradient boosting and gradient boosting with linear Regression. By including logistical regression follow-up time into their model, they achieved an accuracy of 74% for all clinical characteristics. Using a novel machine learning approach, Saqib and et al. [8] evaluated their model's ability to predict 30-day heart failure readmission or death compared to those of other commonly used prediction models. A multi-layer perceptron (MLP) technique has been used to make predictions. Here, area under the curve (AUC) scores of 0.50 to 0.63, Precision–Recall (PR) curve scores of 0.31 to 0.46, and accuracy levels of 64.93 to 76.39% were achieved using machine learning models. Advanced data mining techniques may be utilized to uncover hidden patterns in data, as shown by Milan et al. [9]. Doctors may use the models created by these technologies to make better decisions. These methods are employed in this paper: artificial neural networks, ripper algorithm, and support vector machine. Support vector machine (SVM) is the most accurate algorithm, with an 84% success rate.

Khader et al. [10] have built a machine learning method, which they utilized across a large EMR, to assess the likelihood of heart failure readmissions. The readmission rate was able to obtain up to 16.66% of the variables using just the electronic medical data from a single facility. With existing prediction models for HF readmission rates, the AUCs range from 0.60 to 0.70; with electronic-medical record-wide (EMR-wide) predictive models, they range from 0.78 to 83.19%. In a non-invasive and less expensive procedure, Roohallah et al. [11] used multiple data mining algorithms to predict arterial stenosis. Their predictions were made using the Bagging and C4.5 classification algorithms. The Bagging and C4.5 classification algorithms attained a maximum accuracy of 79.54% and 61.46%, respectively.

Table 1
Comparison of accuracy among the proposed system and existing system.

Sl no	Authors name	Dataset	Algorithms	Accuracy
1	Milan et al. [9]	Cleveland cardiovascular disease dataset of 303 records (14 features)	RIPPER DT ANN SVM	81.08% 79.05% 80.06% 84.12%
2	Khader et al. [10]	Cohort dataset of 1,246 instances (28 features).	NB	83.19%
3	Roohallahet al. [11]	303 individual Rajaie Cardiovascular (34 features)	Bagging C4.5 classification	79.54% 68.96%
4	Sara et al. [12]	1.EnterpriseData Warehouse (EDW) 2.Research Patient Data Repository (RPDR)	LR GB max out networks DUNs	75.4% 73.9% 75.4% 76.4%
5	S.Irin Sherlyet al. [13]	Heart Disease dataset from Switzerland, Hungary,Cleveland and the VA Long Beach 920 clinical incidents (76 Features).	Bagging DT Gradient Boosting DT KNN RF SVM LOR	96.6% 99% 71.6% 98.3% 81.6% 86.6%
6	C. Beulahet al. [14]	Cleveland heart dataset (14 features)	C4.5 RF MP BN NB	79.87% 80.53% 81.52% 84.16% 84.16%
7	Tim et al. [15]	A dataset provided by the Careggi University Hospital, University of Florence, Italy with 2303 instances (38 Features)	RF SVM (linear) SVM (RBF) Boosted trees NN	72% 69% 67% 75% 74%
8	Suveen Angraalet al. [16]	A dataset of 1,767 patients with HFpEF were included from 4 countries	LR (forward selection) LR (lasso regularization) RF GDB SVM (linear kernel) SVM (radial basis)	70.3% 70.3% 76% 72% 72%
9	Our Proposed system	Heart failure clinical records dataset (13 Features)	mANN FDA CIT BTM BLR NN RF SVM RT NB BGLM MARS BoGLM	82.95% 82.95% 79.55% 78.40% 82.95% 84.09% 85.23% 81.82% 81.82% 79.54% 82.95% 84.09% 80.68%

Sara et al. [12] used a long-term healthcare system with computerized medical data to categorize hospitalized heart failure patients. The DUN model achieved the best overall performance after 10-fold cross-validation. 72.4% of the time, the DUNs model is correct. Research by Rashmi et al. [13] used 303, a dataset obtained from Cleveland. 75.55% accuracy was attained using the decision tree approach. Weak classification methods are being improved by C. Beulah et al. [14] who are developing an early detection model for heart disease. Ensemble classification has increased the accuracy of poor classifiers by up to 7%. There were three methods used: the bagging, the hybrid, and the ensemble boosting methods. Using the Bagging Approach, Boosting Approach, Hybrid Approach, and Ensemble Boosting Approach, they were able to achieve maximum accuracy of 80.53%, 75.9%, 85.48%, and 78.8%, respectively, in their experiments. Tim et al. [15] provide a machine learning technique that may improve the classification of SCD, cardiac death, and all-cause mortality risk. Among the categorization methods they have employed are RF, SVM, and ANN. The DUNs model has had a 72.4% success rate. They have a 75% success rate with the boosted trees. In the Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist Research, Suveen Angraal et al. [16] created a model for predicting survival and hospitalizations in outpatients with heart failure (HF) with preserved ejection fraction. Logistic regression with the lasso, RF, gradient descent (GD), and support vector machine are among the five approaches they have used. In this case, logistic regression using a lasso has the best accuracy of 76%. The comparison of model performances is shown in Table 1.

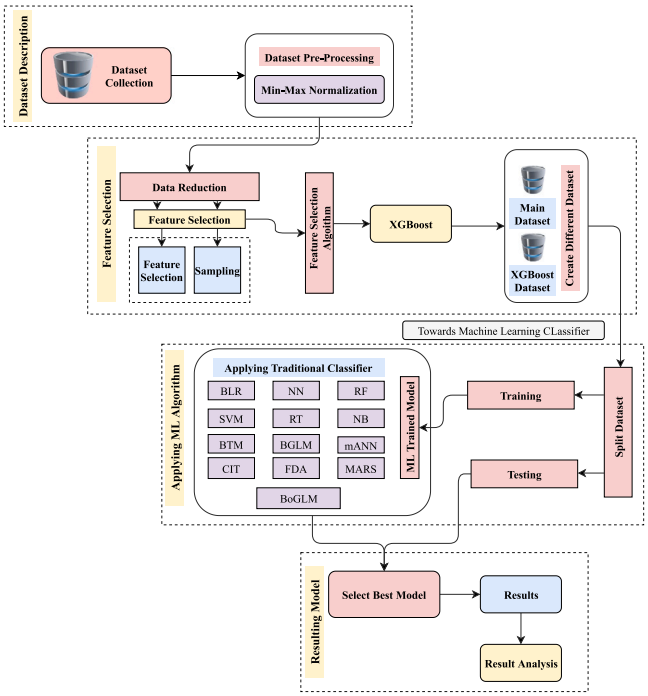


Fig. 1. Diagram of proposed model.

3. Research methodology

This paper provides an outline of how to build a system with machine learning using a chronic heart disease dataset.

3.1. Overview of the proposed model

As part of the data preparation process, one of the initial steps is employing XGBoost selection methods to identify missing values in the dataset. This crucial phase enables the recognition and safeguarding of the most significant and essential attributes for subsequent investigation. The performance of the classifiers is evaluated by utilizing the selected features described earlier in combination with the original characteristics. The objective of this assessment approach is to evaluate the performance of several classifiers in relation to the chosen attributes, offering valuable insights into their skills and constraints in managing the dataset. The workflow depicted in Fig. 1 illustrates the development phase of the proposed model, showcasing the integration of these classifiers into the framework. Each classifier possesses a unique objective, utilizing its expertise to enhance the predictive capabilities of the model. The classifiers in question comprise a heterogeneous range of algorithms, each of which has been meticulously designed utilizing distinct methodologies to efficiently tackle the intricacies inherent in the dataset. The utilization of various classifiers aims to ensure comprehensive and robust coverage in the modeling process, hence enabling a comprehensive and accurate prediction mechanism.

4. Application of the proposed model

Effective application of the suggested framework will contribute to the development of this unique system and address practical issues. The process of the expert system has been illustrated in this section. The working process of the system has been depicted in Fig. 2. This system can help a community health center use the system to identify the heart failure condition of a patient. The processes in Fig. 2 have been shown as follows:

- Step 1: Collected data are uploaded into the database.
- Step 2: Some features will be selected from the dataset for FDA model training
- Step 3: Selected features will be further processed in the train model.
- Step 4: Then output will be generated either “0”or “1”.
  - If it is “0”, that patient is less prone to heart failure
  - If it is “1”, that patient is prone to heart failure.

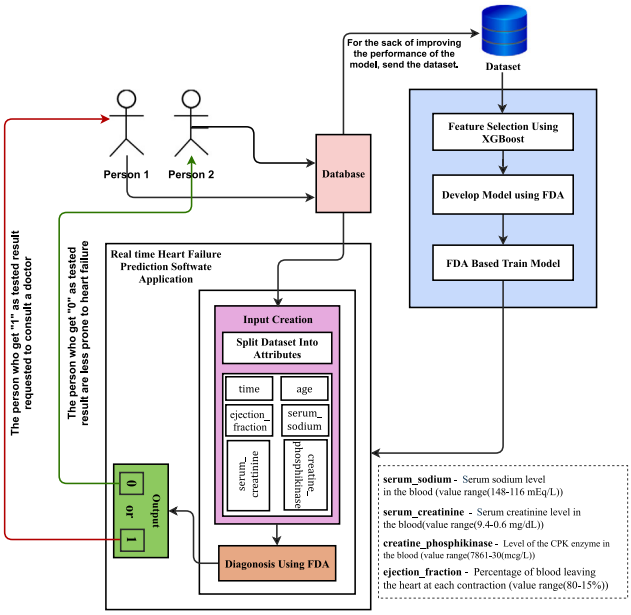


Fig. 2. Application of the proposed model.

- Step 5: If the output is generated as "1", then the patient is suggested to consult a doctor.
- Step 6: To increase the accuracy of the trained model, data is added to the database and utilized to construct a trained model.

5. Implementation

5.1. Dataset description

The dataset employed in this study is obtained from a repository of machine learning records containing clinical heart failure data. The collection comprises 299 instances in total, each distinguished by 13 unique features. These features encompass a multitude of facets pertaining to cardiovascular well-being and the demographics of patients, thereby enhancing the overall comprehension of the dataset. For a detailed overview of the dataset, including statistical measures and descriptive statistics, refer to Table 2.

Data analysts utilize Explanatory Data Analysis (EDA) to evaluate and study large datasets and highlight their main properties. EDA's primary goal is to assist with data analysis before any assumptions are formed. In addition to the many visual EDA methods, there are a few that combine analytical analysis. Fig. 3 shows the EDA analysis of this dataset.

5.2. Data normalization

We preprocessed the dataset using normalization following data collection. Normalization is an indispensable process in the compilation of data. In order to maintain consistency across data features with varying ranges, it is possible to modify numerical columns in a given dataset. One may employ a variety of normalization techniques. Among them is the min–max normalizing technique. It is among the most frequently employed techniques for linearly transforming and normalizing data. The data may be modified to fall within a predetermined range in this instance. The minimum value for each attribute is zero, and the maximum value is one; the remaining values are converted to decimal integers ranging from 0 to 1 [17]. Using Min–Max normalization, the equation is as follows:

$$V' = \left(\frac{V - \min(X)}{\max(X) - \min(X)} \right) * (new_\max(X) - new_\min(X)) + new_\min(X) \tag{1}$$

where,

A is the attribute data,
$\min(X)$ is the lowest absolute measure of X.
$\max(X)$ is the largest absolute measure of X
V' is the new value for every data input.
V is the old value of every data input.
$new_\max(X)$ is the largest measure of the predefined boundary
$new_\min(X)$ is the lowest measure of the pre-defined boundary.

Table 2
Description of dataset.

Feature	Explanation	Range	Measurement
Age	Age of the persons.	40 to 95	Years
Anaemia	Decrease of red blood cells or hemoglobin	0,1 (1:Yes 0:No)	Boolean
creatinine phosphokinase	Level of the CPK enzyme in the blood	23 to 7861	mcg/L
Diabetes	If the patient has diabetes	0,1 (1:Yes 0:No)	Boolean
Ejection fraction	Percentage of blood leaving the heart at each contraction	14 to 80	Percentage
High blood	pressure If a patient has hypertension	0,1 (1:Yes 0:No)	Boolean
Platelets	Platelets in the blood	25.1 to 850	kiloplatelets/mL
Serum creatinine	Level of creatinine in the blood	0.50 to 9.40	mg/dL
Serum sodium	Level of sodium in the blood	113 to 148	mEq/L
Sex	Woman or man	0,1 (1:Yes 0:No)	Binary
Smoking	If the patient smokes	0,1 (1:Yes 0:No)	Boolean
Time	Follow-up period	4 to 285	Days
(Target) death event	If the patient died during the follow-up period	0,1 (1:Yes 0:No)	Boolean

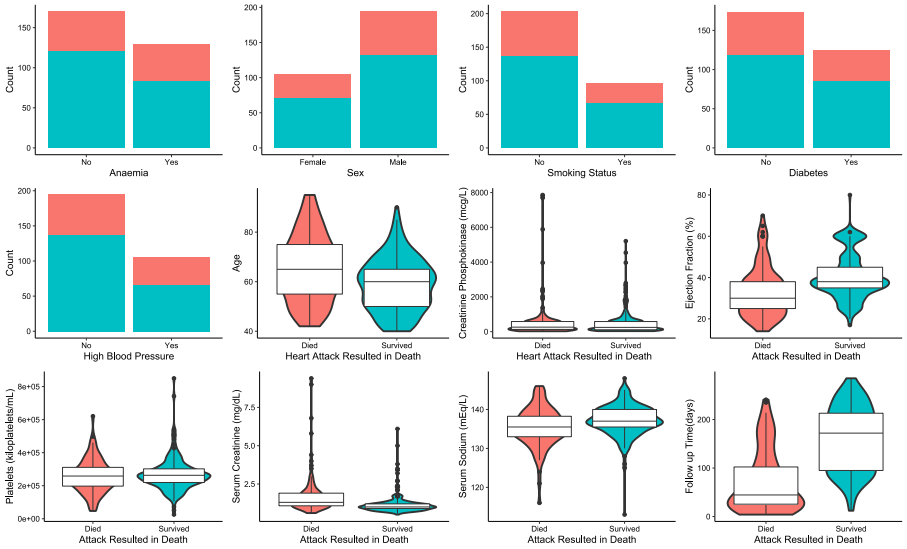


Fig. 3. The histogram of this dataset.

5.3. Correlation

In statistics, a correlation is a plot representing the relationships between two sets of data. How closely two variables are related may be determined by using this method. Values for correlation may range from 0 to 1. If correlation has a positive value, which is $(0 < correlation <= 1)$ then it indicates that the variables are positively correlated, and both of these variables have a tendency to rise in tandem. If 0, then it indicates that there is no relationship between the variables. If correlation has a negative value, which is $(-1 <= correlation < 0)$ then it indicates that the variables have a negative relationship, which means the increase of one variable will decrease another. In the case of a data set containing two sets: $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ the correlation coefficient may be calculated:

$$C = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{2}$$

Here, the sample size is n. $\overline{x}$, and $\overline{y}$ are called the mean values, and ith data values are x_i and y_i [18]. The correlation dataset is shown in Fig. 4.

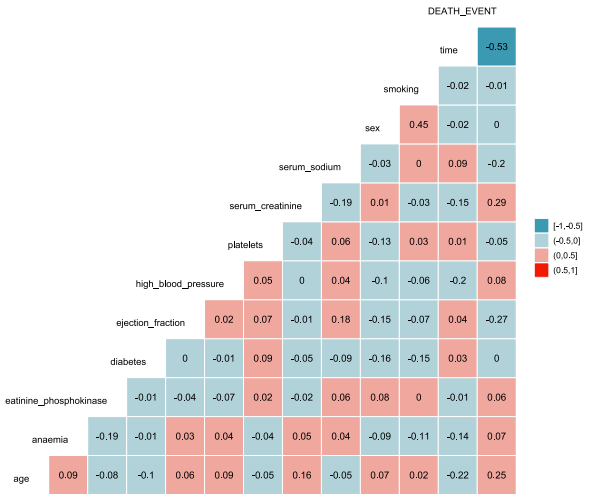


Fig. 4. Result of correlation across all features, with the results being displayed as a heatmap.

5.4. XGBoost feature selection technique

An effective feature selection strategy may enhance the performance of machine learning algorithms by deleting characteristics that are unnecessary, redundant, or unrelated to the job at hand. A trait that has a significant correlation with other features and lowers accuracy is referred to as a redundant one. However, features that do not provide any information about different classes are considered to be irrelevant. An important characteristic that is closer to the root node gets a higher relevance score from a global viewpoint. It is possible to create relevance scores for features using several boosting tree approaches, such as XGBoost. "$Gain$", "$Cover$", and "$Frequency$" [19] are the XGBoost's three alternate feature significance ratings. "$Gain$"refers to the average gain over all splits in which this feature has been employed, while "$Cover$"refers to the average coverage across all splits [20]. Feature significance was rated using the term "benefit"in this research. First, we used the XGBoost algorithm to determine the relevance of each characteristic, and then we compared the results. The XGBoost algorithm assigns each feature a weighted significance value. All of the characteristics are rated based on importance ratings. Using these rated characteristics, we have picked the best possible feature subset.

5.5. Dataset splitting

Machine learning models can only be evaluated if they are tested on a variety of datasets. To prevent overfitting, we divided the dataset into a training set and a test set. If we use the same data size for both the training and testing sets, we will not be able to accurately evaluate the model's performance. To build a more accurate model, we should keep the majority of the data in the training set rather than discarding it. We can eliminate data discrepancies and better understand the model's nature since the training and test sets of data are comparable. When feature selection is done, the dataset is separated into two sections: training and testing. This means that 70% of the data is used for model training, while the remaining 30% is used for model testing.

5.6. Ensemble methods of machine learning

You may train a machine learning model to recognize specific patterns. 13 categorization strategies were employed in our study to train the model. The developed models are then applied to a specific computing problem. It is used to improve the performance of the classification or prediction model and reduce the likelihood that a subpar model would be chosen. It is, in essence, the process of classifying a large amount of data into discrete groups. Labeling or categorizing the classes is a typical practice. Based on a patient's past medical history, machine learning and data mining techniques can be used to forecast the patient's health [21].

5.6.1. Model average neural network (mANN)

Models of nonlinear deep neural networks are adaptable. We can learn map functions for as long as you like. This flexibility results in a final model that is highly changeable. The models mostly depend on training data and random initial weights during training. This issue is resolved by model averaging. Fitting numerous neural network models with the same data set is known as multi-layerANN modeling.

5.6.2. Flexible discriminant analysis (FDA)

FDA can be used to perform Linear Discriminant Analysis (LDA) on generated answers. In order to provide answers for these questions, scores are assigned to the classes in a way that allows regression on X to best predict the revised class labels. The average

squared residual is the discrimination criteria, and it is denoted by

$$ASR = \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n}(\theta_l(g_1) - x_i'\beta_l)^2 \tag{3}$$

where g_i is the class of the ith sample and x_i is the vector of p predictors for the ith sample But FDA is far more than that. Indeed, $x_i'\beta_l$ indicates a linear regression and this may be replaced by non-parametric regression $\eta_l(x_i)$. To overcome the drawbacks of linear separations in LDA, it will be effective. And then, the criterion becomes

$$ASR = \frac{1}{n}\sum_{l=1}^{L}\sum_{i=1}^{n}(\theta_l(g_i) - \eta_l(x_i))^2 \tag{4}$$

5.6.3. Conditional inference tree (CIT)

Conditional Inference Tree, or CIT, is a categorization approach that is based on trees. It is a method of conditional inference for response variables that are multivariate and continuous. The algorithm operates in the following manner:

1. When generating the response variables from a random input, this method looks for a logical null hypothesis. Then, it chooses the input variable with the greatest p-value in relation to the response variable.
2. XGBoost has been used to determine feature relevance in order to avoid overfitting and underfitting.
3. The method divides the input variable into two halves.
4. First and second stages are iteratively executed by the algorithm.

5.6.4. Bagged tree (BT)

The bagging tree approach can lower the DT classifiers' variance. There is a reason why training sample data should be divided into subsets. Random subsets of data can be used to train the decision tree.

5.6.5. Binomial logistic regression (BLR)

A binomial logistic regression forecasts the probability that an observation will belong to one of two groups of a binary dependent variable based on one or more independent factors.

$$p = \frac{1}{(1+e^{-z})} \tag{5}$$

where p is the probability that an event will occur. The probability on an S-shaped curve is between 0 and 1, and z is the linear combination.

5.6.6. Random forest (RF)

A tree-based technique for estimating probability is called random forest. In categorization, the RF is constructed by several decision trees. The amount of trees, their depth, and D and E all play a major role in classification. The impurity quality of the model is enhanced when a decision tree is constructed using CART and the Gini Index. The formula for the Gini Index is:

$$Gini = 1 - \sum_{n=1}^{n} p_n^2 \tag{6}$$

Where p denotes the probability. The Gini index with smaller value indicates less impurity.

5.6.7. Support vector machine (SVM)

A supervised machine learning algorithm is the Support Vector Machine algorithm. Regression and classification work well with it. This approach, which makes use of a decision tree, can form more than two groups. The support-vector, hyperplane, marginal distance, linear and nonlinear separable points, and marginal distance are some of the key words to be aware of when utilizing this approach.

5.6.8. Naive Bayes (NB)

In data mining, Naive Bayes is a popular classification strategy. When there are a large number of parameters, it connects all of the properties. In this approach, Bayes theory is used as the basis. The symbol $p(A \mid B)$ represents a conditional probability, which refers to the likelihood of event A happening, provided that event B has already happened. According to this statement, the joint probability of events A and B is equal to the conditional probability of event A provided that event B has happened, multiplied by the likelihood of event B happening. The algorithm's formula is as follows:

$$p(A \mid B) = \frac{p(B \mid A)p(A)}{p(B)} \tag{7}$$

5.6.9. Multivariate adaptive regression spline (MARS)

To forecast nonlinear outcomes, the multivariate adaptive regression splines approach is employed. The primary benefit of this model is its ability to describe nonlinear connections between independent variables and predictors. To map inputs to a new variable, one can utilize two kernel functions or the value of a parameter. Eqs. (8) and (9) in the context of Multivariate Adaptive Regression deal with the calculation of output values (YY) based on input values (XX) and a threshold value (cc).

$$Y = max(0, X - c) \tag{8}$$

$$Y = max(0, c - X) \tag{9}$$

Where c is the threshold value.

5.6.10. Boosted generalized linear model (BoGLM)

Non-normal error distributions can be taken into consideration via a form of linear regression model called an extended linear model. Random effects are incorporated into the general regression model in extended GLMs. Dependent observations, like repeated observations, can be described with these phrases. The prediction's accuracy may increase with the addition of random variables.

5.6.11. Random tree (RT)

Exact and simplified results for the uniform random categorized tree on n-vertices have been investigated in combinatorial studies. The topic at hand is stochastic asymptotic analysis. Beyond all limits in three ways. There exist only infinite, discrete plants. Fractal tree-like subsets of the infinite-dimensional space l1, which is two-dimensional, are present. The former is restricted, whereas the latter is boundless. Utilizing a straightforward algorithm and mild convergence objections, the finite random tree is generated.

5.6.12. Neural network (NN)

Neural networks, similar to the human brain, search for concealed patterns in massive datasets. Artificial and biological neural networks are capable of performing a variety of neuronal functions. The implementation of neural networks powered by artificial intelligence in trading platforms is expanding at a rapid rate. Neural networks are capable of credit risk assessment, stock price prediction, and security identification. Additionally, custom signals and pricing contracts can be generated.

5.6.13. Bayesian generalized linear model (BGLM)

In order to model data with exponential family responses and covariates, GLMs have been implemented. Frequent illustrations consist of binomial and Poisson response data. The link function delineates the relationship between the model and the observations, while the covariate information framework is established by a regression model. GLM, on the other hand, does not include basic linear models with normal error structures. As is customary in Bayesian analyses of these types of data, the prior must be specified for the regression coefficients. Non-informative priors that are frequently employed consist of diffuse, Jeffrey's, and uniform.

6. Results and discussion

6.1. Feature importance finding

Feature values interpret the significance of having a particular value for a given characteristic in comparison to the forecast we would make if that characteristic had a foundation value. The R programming language was employed to conduct this analysis, and specific packages were utilized to manipulate the dataset. In the dataset, the feature "time"has the highest feature value and the value is 0.042 which means it has a high impact on heart failure. Next is serum_creatinine, its SHAP value is 0.016, the 2nd cause of heart failure between all features. creatinine_phosphokinase and serum_sodium have the same feature value. high_blood_pressure, anaemia, sex, smoking all have the same feature value, 0.001 so these features have the lowest impact on heart failure. Fig. 5 shows the significance of the dataset's features after using XGBoost.

Based on the Gain, Cover, and Frequency values of the dataset, the XGBoost machine learning method is used to identify the most significant features. For "time", Gain has a value of 0.5433, which is the biggest. When "serum_creatinine"is taken into account, Cover has the highest value of 0.2898. "creatinine_phosphokinase"has the highest frequency value of 0.1759. "smoking"has the lowest Gain, Cover, and Frequency values of any characteristic. Our literature will focus on six key variables out of a total of thirteen: "time", "serum_creatinine", "creatinine_phosphokinase", "ejection_fraction", "age", "serum_sodium". These variables were chosen based on the highest Gain, Cover, and Frequency values. We next used machine learning methods to improve the trained model's performance. Table 3 shows the values of Gain, Cover, and Frequency for the dataset's features.

SHAP values assess the significance of having a specific value for a given characteristic in compared to the prediction we had make if that feature had a foundation value. In our used dataset, in term of the feature "time", when the SHAP value is −0.05 to 0.00, "time"is more scattered between 100 to 200. "serum_creatinine", is more distributed between 0.0 to 2.5, when the SHAP value is 0. For "ejection_fraction", it is almost equally distributed for all SHAP value. "creatinine_phosphokinase"is more scattered between 0 to 2000, when SHAP value is −0.02 to 0.02. When the SHAP value is −0.01 to 0.01, "serum_sodium"is more distributed between 130 to 140. While looking at the feature "age", when the SHAP value is 0.00 the "age"is more scattered between 40 to 80.

The SHAP value with XGBoost of selected features is shown in Fig. 6:

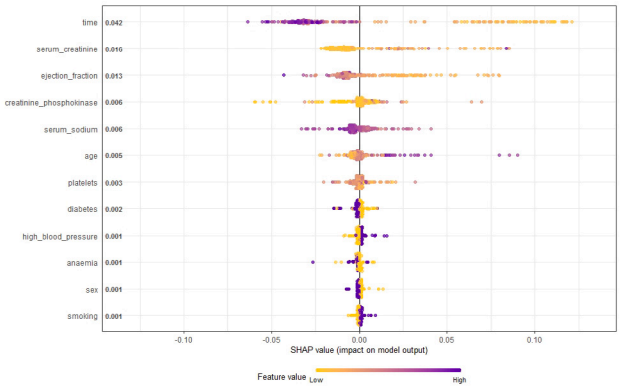


Fig. 5. Feature importance of all features applying XGBoost.

Table 3
Feature Importance Based on Gain, Cover and Frequency Value applying XGBoost.

Feature	Gain	Cover	Frequency
time	0.5402751	0.2156907	0.1443570
serum_creatinine	0.1061804	0.2898113	0.1128609
creatinine_phosphokinase	0.0848320	0.1275443	0.1758530
ejection_fraction	0.0678798	0.1433377	0.1233596
age	0.0550421	0.1086151	0.1653543
serum_sodium	0.0471427	0.0491476	0.0787402
platelets	0.0307467	0.0389418	0.0577428
high_blood_pressure	0.0192308	0.0090655	0.0314961
diabetes	0.0183431	0.0080392	0.0393701
anaemia	0.0170617	0.0046183	0.0472441
sex	0.0100201	0.0024517	0.0131234
smoking	0.0032454	0.0027368	0.0104987

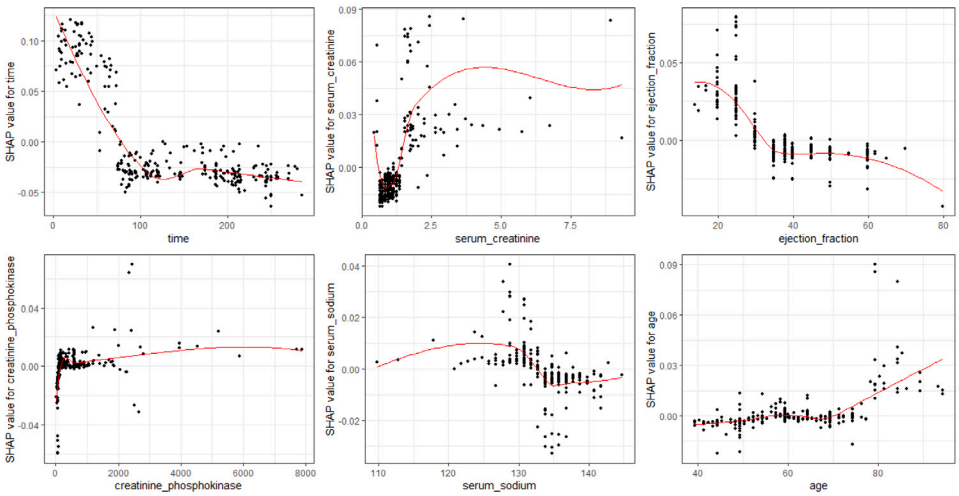


Fig. 6. Selected Features importance on sharp value.

6.2. Clinical references of selected features

serum_sodium: It is essential for heart patients. Patients with heart failure are at significant risk if their serum sodium level is low. Research indicates that patients with heart failure in hospitals have higher morality when their sodium levels are between 135 and 138 mmol/l [22].

serum_creatinine: Patients with heart failure suffer greatly from the rising serum creatinine rate. Twenty to forty percent of patients admitted to hospitals due to heart failure had elevated serum creatinine levels. Higher blood creatinine levels and creatinine

Table 4
Model Performance of all algorithms based Accuracy (%), Sensitivity (%), Specificity (%) and Kappa (%) value.

Algorithms	Full dataset (13 Features)				XGBoost features dataset (6 Features)			
	Accuracy	Sensitivity	Specificity	Kappa	Accuracy	Sensitivity	Specificity	Kappa
mANN	82.95	88.23	81.69	56.12	84.09	81.82	84.85	61.11
FDA	82.95	78.26	84.61	58.75	86.36	80.76	88.71	67.96
CIT	79.55	65.62	87.50	54.59	81.82	71.43	86.67	58.10
BTM	78.40	68.00	82.53	48.77	81.82	71.43	86.67	55.05
BLR	82.95	76.00	85.71	59.56	80.68	70.37	85.25	55.05
NN	84.09	81.82	84.85	61.11	82.95	72.41	88.14	61.08
RF	85.23	77.78	88.52	65.63	81.82	70.00	87.93	58.88
SVM	81.82	75.00	84.38	56.44	81.82	70.00	87.93	58.88
RT	81.82	80.00	82.35	54.63	81.82	80.00	82.35	54.64
NB	79.54	81.25	79.17	46.77	79.55	85.71	78.38	45.60
BGLM	82.95	81.25	89.12	59.56	81.82	73.08	85.48	57.28
MARS	84.09	76.92	87.10	62.62	82.95	72.41	88.14	61.08
BoGLM	80.68	68.97	86.44	55.90	80.68	68.97	86.44	55.90

readings have been associated with longer hospital stays, higher rates of long-term and in-hospital morbidity, and higher readmission rates [23].

creatinine_phosphokinase: Heart failure can be caused with the derangement level of creatinine phosphokinase [24].

ejection_fraction: Heart failure having ejection fraction occurs whenever the left atrium does not circulate blood through the body as well as it should. This affects or harms cardiac muscle and inhibits its capacity to pump blood and thus a patient is affected to heart disease and then heart failure [25].

time: It is hypothesized that aberrant systolic time intervals in human heart failure are caused by a flaw in the mechanical functioning of the heart. With the passage of time, heart disease increase and the risk of heart failure also increases [26]

age: With aging process, both the prevalence and incidence of heart failure (HF) rises. Growing older is a hazard factor for overall morbidity on its own. With the increase of age, the conditions of heart failure patients becoming more critical [27]

6.3. Comparison model performance on different features

Our model is evaluated by different metrics (Accuracy, Specificity, sensitivity, F1 Score, Specificity, Balanced Accuracy, Accuracy Lower, Accuracy Upper, Positive predictive Value, Negative Predictive Value, Prevalence, Detection rate, Detection Prevalence) which is shown in Figs. 7 and 8. Table 4 shows a comparison of the confusion matrix, which we compared. The performance of all models on the full dataset and also the XGBoost dataset is shown in Table 4.

Accuracy is a critical factor to consider when evaluating machine learning algorithms. As previously mentioned, we utilized 13 classifiers. With the XGBoost approach, we specifically selected six out of the 12 attributes. Fig. 7 illustrates the accuracy of various classifiers using XGBoost with all features. While XGBoost achieved the highest accuracy of 86.36% for FDA, the mANN Classifier closely followed with an accuracy of 84.09%. For the full dataset, RF demonstrated the highest accuracy at 85.23%. However, the outcomes surpass all considered metrics. When analyzing just six characteristics using XGBoost, several algorithms achieved the same accuracy of 81.81%. The other models listed above in Table 4 exhibit varying sensitivity levels. As mentioned earlier, the thirteen distinct approaches were applied to the actual dataset rather than the six input characteristics selected by the XGBoost methodology. On our full dataset, we achieved a maximum sensitivity of 0.88 using mANN. The F1 score, calculated using specificity and sensitivity scores, indicated that the FDA methodology had the highest F1-score for both the full dataset and the XGBoost dataset, with a score of 77.78%. For the XGBoost dataset, CIT and BT achieved the same F1-Score of 71.43%. RF on the full dataset yielded an accuracy of 76.36%. Interestingly, having fewer features improved the F1 score. Notably, certain classifiers, including CIT, displayed better F1 scores with six features than with thirteen. Fig. 7 shows how well the model did on the full dataset and the XGBoost feature selection dataset, based on Kappa, sensitivity, specificity, and F1 score.

Specificity is a measure used to determine how many actual negative events are correctly predicted. Models above visualization Table 4 exhibits different specificity compared to those below. RF and mANN yielded the highest specificity of 0.88 when applied to our complete dataset. On the other hand, our XGBoost dataset had a very high specificity of 0.88, which means that it accurately predicted which patients would have a bad outcome from their heart disease when the NN, MARS, and FDA models were applied to the data. With mANN, we achieved the highest balanced accuracy score of 0.84 on the dataset. Averaging NN with FDA resulted in the greatest balanced accuracy score of 0.84 for the XGBoost dataset. However, after feature selection (XGBoost dataset), BLR yielded a lower accuracy score of 0.73, while SVM delivered an even lower score of 0.70. On the dataset, the highest accuracy score of 0.91 was obtained using three models: RF, MARS, and NN. On the XGBoost dataset, using FDA, we achieved an even higher accuracy score of 0.92. For classifiers that achieved a maximum level of accuracy of 0.90, 0.89, 0.89, and 0.87, including BLR, SVM, RT, and NB, the XGBoost dataset presented accuracy scores of 0.88, 0.89, 0.89, and 0.87, respectively. In our comparison of thirteen models, the RF and CIT models demonstrated the highest positive predictive value score of 0.75 on the dataset. Utilizing FDA, MARS, SVM, RF, and NN, we attained the highest positive predictive score of 0.75 on our XGBoost dataset. We compared the negative predictive value of several classifier models, including boosted and regression models, and found that the averaged NN model had the best score of 0.96. For our XGBoost dataset, we utilized RF and SVM to identify a prevalence score value of 0.34 for

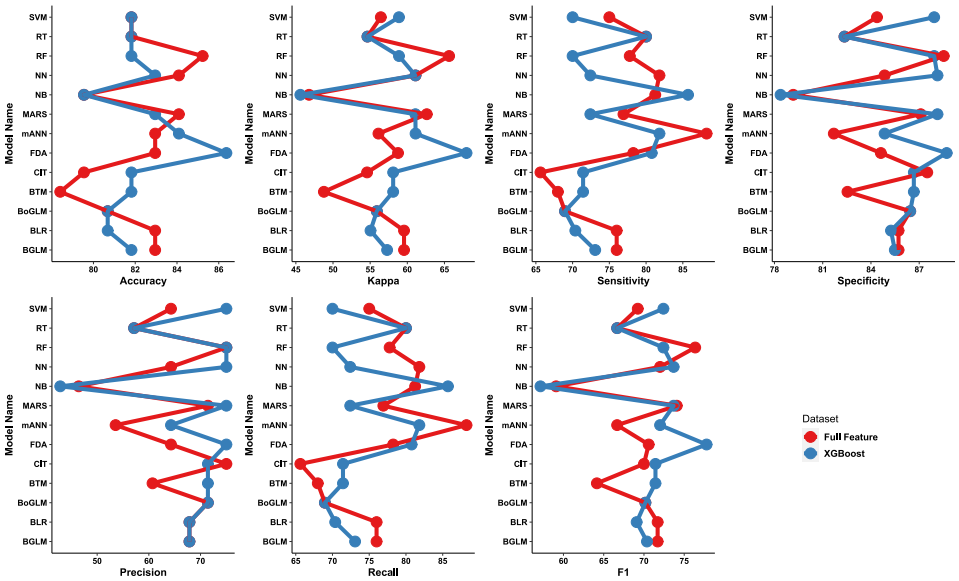


Fig. 7. Model performance on main dataset and XGBoost feature selection dataset based on Accuracy, Kappa, Sensitivity, Specificity, Specificity, sensitivity and F1 score.

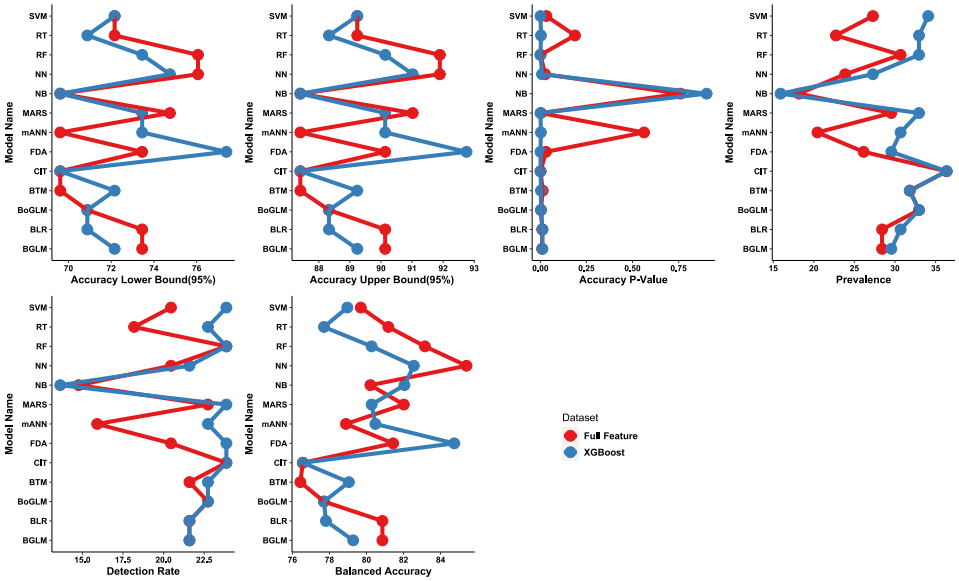


Fig. 8. Model performance on main dataset and XGBoost feature selection dataset based on Accuracy Lower, Accuracy Upper, Accuracy *P*-value, Prevalence, Detection Rate, Balance Accuracy.

heart disease in a specific group, while BLR revealed a prevalence of 0.28. The two datasets exhibited a detection prevalence score of 0.31.

A picture called Fig. 8 shows how well the model did with the main dataset and the XGBoost feature selection dataset based on Accuracy Lower, Accuracy Upper, Accuracy *P*-value, Prevalence, Detection Rate, and Balanced Accuracy.

The maximum TP number is obtained by both CIT and RF in our study, while the highest TN number is obtained by NB on the entire dataset. NB has provided the most inadequate FP value, while RF and CIT have provided the lowest FN values. With regards to the XGBoost dataset, NN, RF, SVM, FDA, and MARS achieve the highest TP number, while NB achieves the highest TN number. Additionally, NB produces the greatest FN and smallest FP values. The FN values produced by NN, RF, SVM, FDA, and MARS are the lowest possible. Table 5 shows a comparison of the confusion matrix, which we compared.

Our model has been developed with thirteen important features. Six of the most notable features are here, obtained through the application of XGBoost's feature selection technique. Using the XGBoost feature selection approach, six distinct and useful features

Table 5
Confusion Matrix of different models.

Algorithm	Full dataset	XGBoost dataset
BLR	$\begin{bmatrix}19 & 6\\9 & 54\end{bmatrix}$	$\begin{bmatrix}19 & 8\\9 & 52\end{bmatrix}$
NN	$\begin{bmatrix}18 & 4\\10 & 15\end{bmatrix}$	$\begin{bmatrix}21 & 8\\7 & 52\end{bmatrix}$
RF	$\begin{bmatrix}21 & 6\\7 & 54\end{bmatrix}$	$\begin{bmatrix}21 & 9\\7 & 51\end{bmatrix}$
SVM	$\begin{bmatrix}18 & 6\\10 & 54\end{bmatrix}$	$\begin{bmatrix}21 & 9\\7 & 51\end{bmatrix}$
RT	$\begin{bmatrix}16 & 4\\12 & 56\end{bmatrix}$	$\begin{bmatrix}16 & 4\\12 & 56\end{bmatrix}$
NB	$\begin{bmatrix}13 & 3\\15 & 57\end{bmatrix}$	$\begin{bmatrix}12 & 2\\16 & 58\end{bmatrix}$
BTM	$\begin{bmatrix}17 & 8\\11 & 52\end{bmatrix}$	$\begin{bmatrix}20 & 8\\8 & 52\end{bmatrix}$
BGLM	$\begin{bmatrix}19 & 6\\9 & 54\end{bmatrix}$	$\begin{bmatrix}19 & 7\\9 & 53\end{bmatrix}$
mANN	$\begin{bmatrix}15 & 2\\13 & 58\end{bmatrix}$	$\begin{bmatrix}18 & 4\\10 & 56\end{bmatrix}$
CIT	$\begin{bmatrix}21 & 11\\7 & 49\end{bmatrix}$	$\begin{bmatrix}20 & 8\\8 & 52\end{bmatrix}$
FDA	$\begin{bmatrix}18 & 5\\10 & 55\end{bmatrix}$	$\begin{bmatrix}21 & 5\\7 & 55\end{bmatrix}$
MARS	$\begin{bmatrix}20 & 6\\8 & 54\end{bmatrix}$	$\begin{bmatrix}21 & 8\\7 & 52\end{bmatrix}$
BoGLM	$\begin{bmatrix}20 & 9\\8 & 51\end{bmatrix}$	$\begin{bmatrix}20 & 9\\8 & 51\end{bmatrix}$

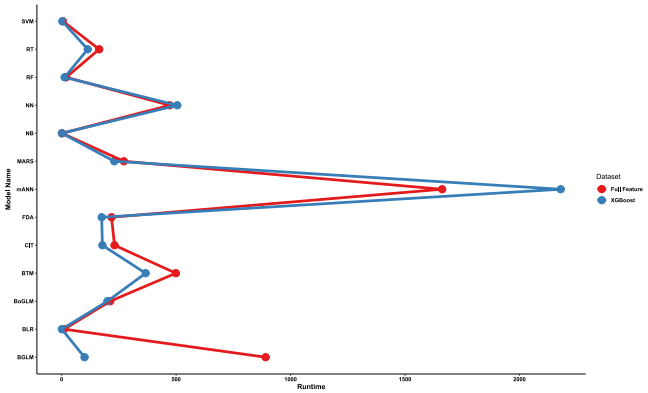


Fig. 9. Diagram of the run time of proposed model.

are chosen, and the SHAP value was also included in the final output. “time”, “creatinine_phosphokinase”, “ejection_fraction”, “serum_sodium”and “age”have been chosen as XGBoost features using the feature selection approach.

The model’s run time is shown in Fig. 9 and has been compared to both the XGBoost and the main dataset run times. It took the smallest amount of time (0.01 s) to run this model on the XGBoost dataset out of all the models tested, and it took the shortest amount of time overall.

6.4. Hyperparameter tuning

The best method for making an algorithm more effective is to tweak its hyperparameters. Hyperparameter tweaking can be used to find an algorithm’s ideal values. Hyperparameter tweaking is a fundamental component of any machine learning algorithms to achieve higher accuracy. As a result, in this work, hyperparameter tuning is carried out to enable the algorithms to produce a more ideal outcome, which is shown in Table 6, and descriptions of all parameters are shown in Table 7. We are able to determine the effective values of all hyperparameters by parameter tuning, and these values are ideal for achieving the best results with this dataset.

Table 6
Used parameters for modeling.

Applied algorithms	Full dataset	XGBoost dataset
mANN	bestTune (size = 5, decay = 0.1, bag = FALSE)	bestTune (size = 1, decay = 0.1, bag = FALSE)
FDA	bestTune (degree = 1, nprune = 8)	bestTune (degree = 1, nprune = 8)
CIT	bestTune (mincriterion = 0.99)	bestTune (mincriterion = 0.99)
BLR	coefficients	coefficients
NN	bestTune (size = 5, decay = 0.1)	bestTune (size = 3, decay = 0.1)
RF	ntree = 500, mtry = 3	ntree = 500, mtry = 2
SVM	cost = 1, degree = 3, gamma = 0.083, epsilon = 0.1, nu = 0.5, rho = −0.24	cost = 1, degree = 3, gamma = 0.17, epsilon = 0.1, nu = 0.5, rho = −0.14
RT	bestTune (cp = 0.022)	bestTune (cp = 0.03)
NB	laplace	laplace
MARS	bestTune (nprune = 8, degree = 1)	bestTune (nprune = 8, degree = 1)
BoGLM	bestTune (mstop = 100, prune = no)	bestTune (mstop = 150, prune = no)

Table 7
Description of classifier's parameters are used for modeling.

Parameters	Description
size	The number of units in the hidden layers as a whole.
decay	Effectively disincentivizes high-weight values in the model.
bag	A classifying mechanism, and bagging is an aggregate learning approach.
degree	The quantity of input data that each piecewise linear function takes into account.
nprune	The number of terms to preserve in the final model.
mincriterion	1 - P-value threshold
ntree	The number of tree of a model. The more tree the more probability to get more accuracy but it takes more memory and time to execute.
cost	Determines how far an SVM should "bend"with the dataset.
gamma	Determines the amount of curvature in a classifier.
epsilon	Establishes a tolerance margin in which errors are not punished.
nu	Limits training errors. This parameter is always between (0,1)].
cp	used to regulate and determine the appropriate tree size.
prune	"prune"is used to reduce overfitting of training dataset.
laplace	A technique for detecting zero-probability scenarios.
mtry	The quantity of factors open for parting at each tree node.
rho	Reduces the impact of previous gradients.
mstop	Boosting iterations

7. Discussion

Using selection techniques, it is possible to increase the number of features selected, as indicated in the figure. As a whole, the mANN (82.95%) and FDA (82.95%) had the best accuracy, while the BTM had the lowest accuracy when all features of an experiment were combined (78.40%). Several important changes occur when the XGBoost technique of selection is applied. FDA had a higher accuracy rate than the BTM and CIT(81.81%), whereas FDT had a lower one (86.36%). Table 1 shows the comparison of our results to existing models and datasets. Davide et al. [7] are the first members of our table to go to work. When working with the same data set, we achieved 83% accuracy in the highest logistic regression, followed by 64% accuracy in RF. Saqib et al. [8] also achieved a high of 76.39% accuracy while working with Western Australian-linked administrative medical data. Tim et al. [15] analyzed data from 2302 patients with hypertrophic cardiomyopathy that included ventricular tachycardia and cardiac failure. Achieved a maximum of 75% accuracy in Boosted tree in their mission. When compared to previous research, our proposed model has great performance, as shown in Table 1.

8. Conclusion

Prior to furthering our investigation, it is imperative to emphasize the criticality of heart failure, a medical condition that results in a substantial annual loss of life. The effectiveness of our predictive models, specifically the FDA model, which achieved a remarkable accuracy of 86.36% in predicting mortality, underscores the potential influence of sophisticated machine learning methods in tackling this urgent public health issue. Due to the critical nature of heart failure, not only do precise prognostications but also prompt interventions are required to mitigate its detrimental effects. Notwithstanding the encouraging outcomes identified in our examination, it is critical to recognize the constraints imposed by the comparatively limited dataset at our disposal. The aforementioned limitation necessitates the collection of more extensive and varied datasets in order to augment the resilience and applicability of our predictive models. Furthermore, it is recommended that forthcoming investigations delve into the incorporation of advanced feature engineering methods and machine learning algorithms in order to enhance and refine the precision of mortality prognostications pertaining to instances of heart failure.

It is expected that the results of our proactive strategy will transcend mere statistical significance. Our objective is to achieve a concrete decrease in heart failure-related mortality rates by promptly identifying and intervening in individuals who are at risk. The potential transformative impact of incorporating our predictive models into clinical practice lies in the provision of healthcare professionals with invaluable insights regarding the probability of mortality in cases of heart failure. By adopting this proactive approach, the organization is making a positive contribution to the overall objective of improving public health outcomes, thereby fostering a healthier and more resilient populace.

CRediT authorship contribution statement

Md. Mehedi Hassan: Conceptualization, Methodology, Software, Visualization, Writing – original draft. **Sadika Zaman:** Methodology, Software, Visualization, Writing – review & editing. **Md. Mushfiqur Rahman:** Data curation, Writing – review & editing. **Anupam Kumar Bairagi:** Writing – review & editing. **Walid El-Shafai:** Writing – review & editing. **Rajkumar Singh Rathore:** Writing – review & editing. **Deepak Gupta:** Writing – review & editing.

Declaration of competing interest

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript

Data availability

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

Acknowledgments

The authors are very grateful to all the institutions in the affiliation list for successfully performing this research work. The authors would like to thank Prince Sultan University, Saudi Arabia for their support.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Emmons-Bell S, Johnson C, Roth G. Prevalence, incidence and survival of heart failure: a systematic review. Heart 2022;108(17):1351–60.
[2] Regitz-Zagrosek V, Gebhard C. Gender medicine: effects of sex and gender on cardiovascular disease manifestation and outcomes. Nat Rev Cardiol 2023;20(4):236–47.
[3] Tan Y, Duan X, Wang B, Liu X, Zhan Z. Murine neonatal cardiac B cells promote cardiomyocyte proliferation and heart regeneration. NPJ Regener Med 2023;8(1):7.
[4] Hasan N, Bao Y. Comparing different feature selection algorithms for cardiovascular disease prediction. Health Technol 2021;11:49–62.
[5] Yekkala I, Dixit S. Prediction of heart disease using random forest and rough set based feature selection. Int J Big Data Anal Healthc (IJBDAH) 2018;3(1):1–12.
[6] Latha C, Jeeva S. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. In: Informatics in medicine unlocked, vol. 16, Elsevier; 2019.
[7] Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Med Inform Decis Making 2020;20(1):1–16.
[8] Awan SE, Bennamoun M, Sohel F, Sanfilippo FM, Dwivedi G. Machine learning-based prediction of heart failure readmission or death: implications of choosing the right model and the right metrics. ESC Heart Failure 2019;6(2):428–35.
[9] Kumari M, Godara S. Comparative study of data mining classification methods in cardiovascular disease prediction 1. Citeseer; 2011.
[10] Shameer K, Johnson KW, Yahi A, Miotto R, Li L, Ricks D, et al. Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: a case-study using mount sinai heart failure cohort. In: Pacific symposium on biocomputing 2017. World Scientific; 2017, p. 276–87.
[11] Alizadehsani R, Habibi J, Sani ZA, Mashayekhi H, Boghrati R, Ghandeharioun A, et al. Diagnosing coronary artery disease via data mining algorithms by considering laboratory and echocardiography features. Res Cardiovas Med 2013;2(3):133.
[12] Etu E-EE, Aguwa C, Monplaisir L. The impact of machine learning algorithms on benchmarking process in healthcare service delivery. IEOM Soc Int 2018;1229–30.
[13] Sherly SI, Mathivanan G. An ensemble basedheart disease predictionusing gradient boosting decision tree. Turk J Comput Math Educ (TURCOMAT) 2021;12(10):3648–60.
[14] Latha B, Christalin C, Carolin Jeeva S. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. In: Informatics in medicine unlocked, vol. 9, Elsevier; 2019, 16: 1.
[15] Smole T, Žunkovič B, Pičulin M, Kokalj E, Robnik-Šikonja M, Kukar M, et al. A machine learning-based risk stratification model for ventricular tachycardia and heart failure in hypertrophic cardiomyopathy. Comput Biol Med 2021;135:104648.
[16] Angraal S, Mortazavi BJ, Gupta A, Khera R, Ahmad T, Desai NR, et al. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. JACC: Heart Failure 2020;8(1):12–21.
[17] Swift D, Cresswell K, Johnson R, Stilianoudakis S, Wei X. A review of normalization and differential abundance methods for microbiome counts data. Wiley Interdiscip Rev Comput Stat 2023;15(1):e1586.
[18] Bruno M, Sommer R. On fits to correlated and auto-correlated data. Comput Phys Comm 2023;285:108643.
[19] Guo J, Yang L, Bie R, Yu J, Gao Y, Shen Y, Kos A. An XGBoost-based physical fitness evaluation model using advanced feature selection and Bayesian hyper-parameter optimization for wearable running monitoring. Comput Netw 2019;151:166–80.
[20] Chakraborty S, Bhattacharya S. Application of XGBoost algorithm as a predictive tool in a CNC turning process. Rep Mech Eng 2021;2(1):190–201.
[21] Hassan MM, Mollick S, Yasmin F. An unsupervised cluster-based feature grouping model for early diabetes detection. Healthc Anal 2022;2:100112.
[22] Weissler AM, Harris WS, Schoenfeld CD. Systolic time intervals in heart failure in man. Circulation 1968;37(2):149–59.
[23] Metra M, Cotter G, Gheorghiade M, Dei Cas L, Voors AA. The role of the kidney in heart failure. Eur Heart J 2012;33(17):2135–42.
[24] Fukuta H, Little WC. Contribution of systolic and diastolic abnormalities to heart failure with a normal and a reduced ejection fraction. Prog Cardiovasc Dis 2007;49(4):229–40.
[25] Evans T, Stewart R. Should beta-blockers be recommended after myocardial infarction when left ventricular ejection fraction is normal? Heart 2023;109(15):1128–9.
[26] Abraham WT. Managing hyponatraemia in heart failure. US Cardiol 2008;5(1):5.
[27] Rich MW, McSherry F, Williford WO, Yusuf S, Digitalis Investigation Group. Effect of age on mortality, hospitalizations and response to digoxin in patients with heart failure: the DIG study. J Am Coll Cardiol 2001;38(3):806–13.

Md. Mehedi Hassan, an IEEE Member, is pursuing an M.Sc. in computer science and engineering at Khulna University. As Founder and CEO of Virtual BD IT Firm and VRD Research Laboratory, he is a respected leader in biomedical engineering and data science. Proficient in association rule mining, he holds several patents, contributing significantly to scientific advancement.

Sadika Zaman, a computer science and engineering undergraduate at North Western University, Khulna, Bangladesh, is dedicated to addressing social issues and making a positive impact on society. As a budding researcher, she has contributed significantly to biomedical engineering, publishing nine articles in esteemed international journals and conferences.

Md. Mushfiqur Rahman, a graduate of the University of Dhaka with a B.S. in Statistics, combines statistical expertise with a passion for machine learning. Eager to apply statistical methods to real-world challenges, he is particularly interested in leveraging analytical skills for clinical research.

Anupam Kumar Bairagi, Senior Member of IEEE, holds B.Sc. and M.Sc. degrees in computer science and engineering from Khulna University (KU), Bangladesh, and a Ph.D. in computer engineering from Kyung Hee University, South Korea. As a Professor at KU, he has authored around 60 publications in wireless resource management, healthcare, IIoT, cooperative communication, and game theory.

Walid El-Shafai received B.Sc., M.Sc., and Ph.D. degrees in Electronics and Electrical Communication Engineering. Since January 2021, he has been a senior researcher/Assistant Professor with the Security Engineering Laboratory at Prince Sultan University, Riyadh, Saudi Arabia. He is also an Assistant Professor at the Faculty of Electronic Engineering, Menoufia University. His research interests include Computer vision, AI, and cybersecurity applications.

Rajkumar Singh Rathore is working as Programme Director for M.Sc. Information Technology Management and Technology Project Management in the Department of Computer Science at Cardiff Metropolitan University's School of Technologies, United Kingdom. He is the Fellow of HEA UK and Senior Member of IEEE. He has several years of rich experience in quality of teaching, learning and research excellence.

Deepak Gupta received a B.Tech. Degree in 2006 from the Guru Gobind Singh Indraprastha University, Delhi, India. He received an M.E. degree in 2010 from Delhi Technological University, India, and Ph.D. degree in 2017 from Dr. APJ Abdul Kalam Technical University (AKTU), Lucknow, India. He completed his Post-Doc from the National Institute of Telecommunications (Inatel), Brazil, in 2018.

Environmental Science and Pollution Research
https://doi.org/10.1007/s11356-024-32555-z

RESEARCH ARTICLE

Emission performance evaluation of urban passenger car fleet through field investigation data in a megacity

Abhinav Pandey[1] · **Govind Pandey**[2] · **Rajeev Kumar Mishra**[3]

Received: 17 April 2023 / Accepted: 15 February 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Continued improvements in living standards and the economic well-being in the megacities have led to a huge surge in vehicular density. The worst environmental outcome of the same has been persistent unsafe urban air quality, thanks to vehicular emission. Further, the existing inspection and maintenance programs, conceived to check such emission remain largely ineffective, particularly in developing countries. This is due to the absence of a thorough assessment of the vehicle's compliance with the in-use emission norms generated through reliable field investigation data. To address this gap, the present comprehensive study collected real-time tailpipe emission data from 2040 cars in Delhi, India. Exhaust emission parameters, namely, CO (carbon monoxide), HC (hydrocarbon), and SE (smoke emission), were recorded from both petrol and diesel-driven cars of private ownership, in collaboration with the emission compliance test centers. The performance of cars was assessed in terms of their compliance with the in-use BS (Bharat Stage) emission norms. The one-of-its-kind study reported the petrol cars to be highly compliant toward the BS IV norm while faring even better toward BS II for both CO and HC emissions (80–90%). The conformance to the HC norm was found to be typically better than that for CO (85–90% versus 75–80%). For the diesel-driven cars, BS III compliance levels were reported relatively better compared to BS IV (90% in the case of the former against 80% in the latter's case). Further, the study puts forward a clear indication that the in-use emission norm and maintenance status of cars have a direct and negative relationship with tailpipe emission parameters. Cars of both overseas and domestic origin have almost equal degrees of compliance with the emission norms (over 80% in any case). The study recommends the incorporation of these two critical vehicular variables, i.e., maintenance status and in-use emission standard in the emission certification policy.

Keywords Emission compliance · Vehicular density · Maintenance status · Car emission · Emission policy · Emission norm

Introduction

In the preceding few decades, most countries have witnessed an improved and sustained socio-economic scenario, including developing nations in South-East Asia. This has led to the transformation of the city areas into urban conglomerations. Such urbanization, over the years, has resulted in a huge surge in the vehicular population and the perennially poor urban air quality, attributable to the tailpipe emissions, construction activities, re-suspension of road dust, etc. (Kumar et al. 2017; Gulia et al. 2018). A review of vehicle numbers in terms of the sales data for passenger cars worldwide shows a substantially higher growth rate of 21.16% (in 2019 compared to 2010) and 17.97% (in 2019 against that in 2005). While 45.25 million numbers of such cars were old globally in 2005, 2019 witnessed 90.42 million units, thereby achieving the unwanted distinction of twice the figures in less than a 15-year timeframe (OICA 2019).

The deteriorating ambient air quality has triggered over 7 million casualties annually. This excludes, but is not limited to, the numerous cases of respiratory ailments, morbidity, and mortality (WHO 2019). Further, exposures to vehicular emissions (VE) are now known to trigger numerous serious health conditions *inter alia*, cardiovascular disease(es),

Responsible Editor: Philippe Garrigues

✉ Rajeev Kumar Mishra
rajeevkumarmishra@dtu.ac.in

[1] Jacobs Engineering Group Inc., Gurugram 122 016, India

[2] Civil Engineering Department, Madan Mohan Malaviya University of Technology, Gorakhpur 273 010, India

[3] Department of Environmental Engineering, Delhi Technological University, Delhi 110 042, India

Published online: 23 February 2024

weakened pulmonary function, carcinogenicity, and deaths in severe cases (Ogunseye et al. 2018; Rice et al. 2015; Wong et al. 2019). Evidently, the criteria air pollutants, CO (carbon monoxide) and HC (hydrocarbons), are the two major entities that the vehicular sector contributes to (about 90% of total CO and HC emissions). Further, among the passenger cars section, the light-duty vehicles (LDVs—both petrol or gasoline and diesel-driven), constituting a large segment of all vehicle types, are the foremost contributors to the transportation-generated CO and HC emissions (Bella and Venkateswar 2010; Lyu et al. 2020; Fernandes et al. 2022). Recent studies point out that over 70–80% of urban air pollution in developing countries is credited to automobile emissions triggered by an abundant number of aged passenger cars with poor inspection and maintenance records (Karthikeyan and Prathima 2016; Sharma and Marechal 2019). This scenario is further worsened by a lack of adequate surface transport infrastructure and acceptable fuel quality (Wang et al. 2010; Lyu et al. 2020).

It is because of the substantial number of in-use vehicles that the focus of automobile emission control has been increasingly shifting from newer to aged vehicles (Zhong et al. 2022). Although at-the-source control of exhaust emission from vehicles is one of the most promising measures for pollution mitigation, it calls for a very stern implementation and monitoring system, policy revisits and upgrades, and public awareness (Wang et al. 2014; Gulia et al. 2020 and Pandey et al. 2022a, 2022b). Comparing the scenarios of emission certification policy in India versus the U.S.A., the latter's PCVEs (Programs to Control Vehicle Emissions) appear to be more effective because of the technical and legal framework put in place by the Environmental Protection Agency (USEPA). Such an intricate mechanism is successful in the implementation and compliance of the pertinent policies (Bandivadekar et al. 2015; Raparthi and Phuleria 2022). As a result of stringent test criteria, the pass percentage of LDVs through the designated laboratories is as low as 15% (He et al. 2017).

Comparing India, China, and Brazil on the I/M structure front, in China, the existing inspection and maintenance system for the in-use vehicle is being improved to detect and remove (off-road) older motor vehicles (Nesamani 2010; Lyu et al. 2020). The policy regulators in Brazil, having observed relatively stronger constraints regarding the in-use LDV's emission conformance programs, are attempting to improve their I/M structure too (Dallman 2020; Ribeiro et al. 2021). The accurate evaluation of in-use LDVs for their actual emission is a challenging task in the absence of techno-commercial support in many parts of the world, especially in developing countries (Ventura et al. 2020; Maxwell and Hannon 2017). Another peculiarity lies in the fact that apart from having more efficiency in the compliance by the new passenger cars, the in-use vehicles also are subject to inspection, repair, maintenance, and durability requirements prescribed under US PCVE (Mock and German 2015; Rodríguez et al. 2019).

European Union has a more structured and effective framework for emission testing of in-use vehicles, such as CoP (conformity of production), ISC (in-service conformity), and ISV (in-service verification). This robust mechanism makes OEMs an active stakeholder in emission testing, results, and corrective and/or preventive actions, although, recently, it has been developing various elements and procedures for improved efficiency of overall testing (EU 2023).

The existing environmental policy and concurrent regulation (dictating the I/M program) in India makes it mandatory for all in-use cars (regardless of country of origin, body type, weight, etc.) to possess a valid Pollution Under Control (PUC) certificate.

This PUC certificate (not to be confused with the fitness certificate) requires the vehicles to undergo emission testing at government-authorized testing centers at least annually. The compliance certificate is issued only in case the vehicle complies with the limits of tailpipe CO, HC, and SE prescribed by the in-use emission norms (CPCB 2010; SIAM 2018; Dandapat et al. 2020). This system at present has various grey areas and shortcomings making it largely ineffective. Most issues associated with the policy can be grouped into technical, infrastructure-related, and societal categories (Fig. SM3).

From the literature review on the status of emission recertification policy, it is revealed that such policies for in-use passenger cars in most developing nations are of a very basic nature when compared to the U.S.A., E.U., U.K., Brazil, Germany, and China. For example, these countries investigate tailpipe emission compliance in loaded and real-world driving conditions; the comprehensive field investigation of actual emission compliance levels of the light-duty cars to the corresponding emission standards (especially across a model-wise, more representative dataset) is rarely available; the emission conformance is checked against the norm during the manufacturing of the car and not the latest in-use standard. This scenario too needs to be explored in terms of how emission compliance might vary with stringency or progression of such standards; the phasing-out policy is only vehicle age-specific, and the other important independent variables are not considered, for example, the maintenance aspect and the applicable emission norm; emission testing infrastructure including instrumentation and data QA/QC (quality assurance/quality control) is an area of huge concern; linkage of the emission data to a centralized system of data storage and retrieval is missing. This situation does not let the gross emitters be screened and identified, and there is hardly any accountability or involvement of the OEMs in the policy (post-sales).

Therefore, the present research investigated the 2040 numbers of both petrol and diesel-driven, privately owned light-duty cars for their tailpipe parameters (CO and HC for petrol-driven passenger cars (PDPCs) and SE for diesel-driven passenger cars (DDPCs)). The overall objective of the research is outlined below:

- Collection of the real-time emission data from the cars upon their appearance at the PUC certification centers across Delhi, India
- Analysis of the data for three dataset scenarios, the entire top 3/top 5 and top 16/top 13 models of PDPPs and DDPCs, respectively, and assessment of the compliance levels of the cars to the relevant emission norms,
- Evaluation of the effect of the vehicle's maintenance category and the emission norm (at manufacturing time) on the car emission

The comprehensive testing program comprised nine numbers of emission monitoring stations (referred to as PUC centers) located across Delhi (Table SM1). The program lasted for about 11 months collecting real-time exhaust emission parameters, CO, and HC in idling and/or fast idling test conditions and SE in FAST mode. The study presents a first-of-a-kind in situ exploratory assessment of emission compliance levels of passenger cars thereby drawing useful inferences for the betterment of the emission recertification policy in the country.

Materials and methods

Study location and tailpipe emission data collection

The in situ testing of PDPCs and DDPCs was conducted in association with the PUC centers located across respective Regional Transport Offices in various districts of the National Capital Territory (NCT) of Delhi, India (Fig. SM1, Tables SM1 and SM2). The vehicle testing program attended to 2040 passenger cars (n = 1580 for PDPCs and n = 460 for DDPCs). As part of the prevalent emission compliance certification policy in Delhi, India, the PDPCs are tested for their emission (CO and HC in idle and fast idle conditions both while CO, HC, CO_2, O_2, and λ in the fast idle condition), and the DDPCs are checked for their smoke emission (SE) in FAST (free-acceleration smoke test) in accordance with the criteria prescribed by the Ministry of Road Transport and Highways, Government of India (MoRTH 2020).

The in-use emission norm applicable during the manufacturing year of the car is then considered, and the PUC certificate is issued to them accordingly (upon conformance) with a particular validity period. As the certificate is issued only based on CO, HC, and SE values in the tailpipe emission for PDPCs and the DDPCs respectively, the present study also used the same parameters in the assessment of compliance. Subsequently, the cars reporting for their exhaust emission certification at the test (PUC) center were halted at the assessment platform and investigated for exhaust emission (CO, HC, and SE in the idle, fast idle, and FAST test modes). Also recorded were car and engine-related independent variables.

The independent car variable parameters included make, model, in-use emission standard as applicable during manufacturing, and the car's maintenance status (in terms of very good/good/poor/unsatisfactory). The dependent variables included tailpipe emission parameters, e.g., CO and HC for PDPCs in both idle and fast idle test modes and SE (in HSU) for DDPCs in the FAST mode. This sample fleet size (n = 2040) was selected so that it appeared to represent the maximum makes and models of the passenger cars plying in Delhi. To ascertain the car models for the present study without any biases, the car sales data and market share over the years were collected (Team BHP 2022; OICA 2019) and analyzed which shows that:

- MSIL (Maruti Suzuki India Limited) has been holding the highest market share (40–50%). In the present study, we have tested the maximum number of MSIL cars (N = 735; 46.52% for PDPCs and N = 155; 33.70% for DDPCs) accordingly.
- HMIL (Hyundai Motors India Limited) has been the second-highest car seller in India with about 15–17% of total sales attributed to it. HMIL was the second-highest number of OEMs tested in the present study (N = 381; 24.12% for PDPCs and N = 57; 12.40% for DDPCs).
- HCIL (Honda Cars India Limited) commanded the third position in case of PDPC sales until 2019, whereas 2020 saw a major decline in numbers. However, its proportion in in-use petrol cars still remains significant, and accordingly, we tested HCIL cars (N = 205; 12.97%). In the case of DDPCs, the third spot is closely occupied by at least three manufacturers, i.e., MML, TKMPL, and FIPL (Mahindra Motors Limited, Toyota Kirloskar Motors Private Limited, and Ford India Private Limited respectively). The present study followed such a trend in testing the number subsequently (MML—54, 11.74%; TKMPL—46, 10%; and FIPL—38, 8.26% respectively).

In model terms, we tested the following numbers which were consistent with the above explanation:

- PDPCs—Alto (n = 133)/Amaze (n = 65)/Baleno (n = 90)/Celerio (n = 35)/City (n = 85)/Dzire (n = 110)/Eeco (n = 9)/Grand i10 (n = 55)/i10 (n = 107)/i20 (n = 96)/Jazz (n = 21)/Nexon (n = 6)/Santro (n = 33)/Swift (n = 147)/Verna (n = 17)/Wagon-R (n = 106)

– DDPCs—Amaze ($n = 19$)/Brezza ($n = 26$)/Ciaz ($n = 18$)/Duster ($n = 16$)/Dzire ($n = 37$)/Ecosport ($n = 21$)/Figo ($n = 15$)/Fortuner ($n = 15$)/i20 ($n = 16$)/Scorpio ($n = 20$)/Verna ($n = 19$)/XUV 500 ($n = 23$)

The snapshot of the market share and sales data for the passenger cars may be referred from Annexure 1 and 2 of the SM. For the ease of data analysis and based on the proportion in the overall sample, 6 datasets were ascertained, namely, entire top 3 makes and top 16 models for PDPCs (whole dataset, $N = 1580$; top 3 makes, $n1 = 1321$; % of $N = 83.60$; top 16 models, $n2 = 1115$; % of $N = 70.57$) and entire top 5 makes and top 13 models for DDPCs (whole dataset, $N = 460$; top 5 makes, $n1 = 350$; % of $N = 76.08$; top 13 models, $n2 = 293$; % of $N = 63.70$).

The overall methodology is presented in Fig. SM2, entailing the research workflow.

Sample (vehicles) characteristics

Make-wise distribution of both PDPCs and DDPCs is provided in Tables SM1 and SM2. Following the frequency analysis, MSIL, HMIL, and HCIL fared as the top 3 makes, while Alto, Amaze, Baleno, Celerio, City, Dzire, Eeco, Grand i10, i10, i20, Jazz, Nexon, Santro, Swift, Verna, and Wagon-R stood as top 16 models for PDPCs having represented over 70% of the total numbers ($n = 1580$). Similarly, for the DDPCs, TKMPL, HMIL, MML, MSIL, and FIPL fared as top 5 makes whereas Swift, i20, XUV 500, Dzire, EcoSport, Amaze, Verna, Duster, Fortuner, Brezza, Figo, Ciaz, and Scorpio were found as 13 models constituting over 60% of the total ($n = 460$). Both the PDPC and DDPC fleets had almost equal proportions of the domestic and overseas car manufacturers of varying characteristics (Tables SM3).

Field instrument characteristics

A four (4)-gas analyzer (model number Oz-Gas-04) and a smoke meter (model no. Oz-Smoke-02) of the Ozone make were used to record the mass concentrations of tailpipe exhaust constituents from PDPCs and DDPCs respectively under the prescribed test modes. The particulars of both testing instruments in terms of their measurement type, range, resolution, operating principle, and other characteristics are outlined in Table SM4. Further, the accuracy levels of the instruments are outlined as Oz-Gas-04—CO = + 1%; HC = + 1% (petrol cars/PDPCs); and Oz-Smoke-02—HSU = + 1% (diesel cars/DDPCs). The type test approval showing compliances of both the instruments toward the manufacturing and performance criteria as specified by the MoRTH was verified with the government and OEM (original equipment manufacturer) relevant web portals respectively. The "Data analysis packages and tools" section of the SM may be referred to for further details.

Data analysis packages and tools

Bearing in mind the size and complexity of the data, the input variables and dataset scenarios, and the study objectives, the present work used the IBM SPSS package (version 23) and Origin package (version 2021) for the tailpipe emission data processing and analysis. In general, the descriptive and inferential statistical techniques were used, for example, the Box and Whisker and/or Box plots to determine statistically significant relationships, if any, between the compliance levels, applicable emission norm, and maintenance category of the petrol and diesel-driven passenger cars plying on roads of the megacity of Delhi ($N = 2040$).

Results and discussion

Vehicle and engine characteristics

The field-assessed characteristics across all the makes and models of both PDPCs and DDPCs are presented in Table SM5. As is evident, the vehicles covered in the present study have a good deal of heterogeneity for almost all vehicle and engine-related parameters recorded (vehicle age, mileage, kerb weight, engine power, torque, compression ratio, bore*stroke, number of cylinders, etc.).

Analysis of emission (recertification) compliance levels

The compliance status to the existing emission norms (as part of the concurrent environmental policy), in terms of the percentage of the total number of petrol and diesel-driven passenger cars vehicles that underwent this study for different datasets, is presented in Figs. 2, 3, 4, 5, 6, and 7. All three dataset scenarios were used, i.e., entire top 3 makes and top 16 models for petrol-driven passenger cars and entire top 5 makes and top 13 models for diesel-driven passenger cars respectively.

The compliance levels were checked for two Indian emission norms, such as BS II and BS IV (in-use) for petrol cars and BS III and BS IV for diesel cars. In line with the emission standards, petrol cars were assessed for compliance with CO in both idling and fast idling conditions (BS IV), CO only in idling mode (BS II), and HC also only in idling mode (BS IV). On the other hand, diesel cars were analyzed for their compliance levels in terms of HSU in FAST mode (for both BS norms) (Table 1).

All dataset scenarios of petrol cars depicted very high levels of compliance toward the BS IV norm (the latest and

Table 1 Tailpipe emission norms used in the present study for vehicle compliance

PDPCs			
Emission standard	CO during idling (%)	CO during Fast idling	HC during idling
BS IV	0.3	0.2	200
BS II	0.5	Not defined	
DDPCs			
Emission standard	HSU during FAS Test		
BS IV	50		
BS III	65		

Source—SIAM 2018; MoRTH 2020

the most stringent at the time of the present tailpipe emission testing program) and even better toward BS II for CO and HC emissions. The reason is the wider range of the BS II (older) norm which lets a higher number of cars pass through (Figs. 1 and 2).

Model-wise compliance status finds Amaze, Baleno, Eeco, Grand i10, and Nexon to be the most compliant models for various test parameters achieving almost 100% compliance toward BS IV norms. A few more models get added to the best-complying models in CO @ Fast idling conditions (Fig. 3).

Although the observed prominent level of compliance is largely attributable to the improvement in engine designs in the recent decade, however, it is also due to the leniency of the applicable emission norms. The wider range of CO and HC concentrations both allows the vehicles to pass through the PUC certification and keep plying on the road even though it may be compliant just by a parts per million or a decimal fraction of percent concentration (in the case of HC and CO emissions) still emitting environmentally worse values in terms of mass emission. The leniency in the PUC (in-use) certification norm is drawing criticism (Sharma et al. 2020) and should be improved for stringency even in fast idling test modes.

The status of emission compliance in relation to the diesel-driven cars for all the dataset scenarios is produced in Fig. 4, accompanied by Figs. 5 and 6 representing the same for a car manufacturer and in terms of car models, respectively.

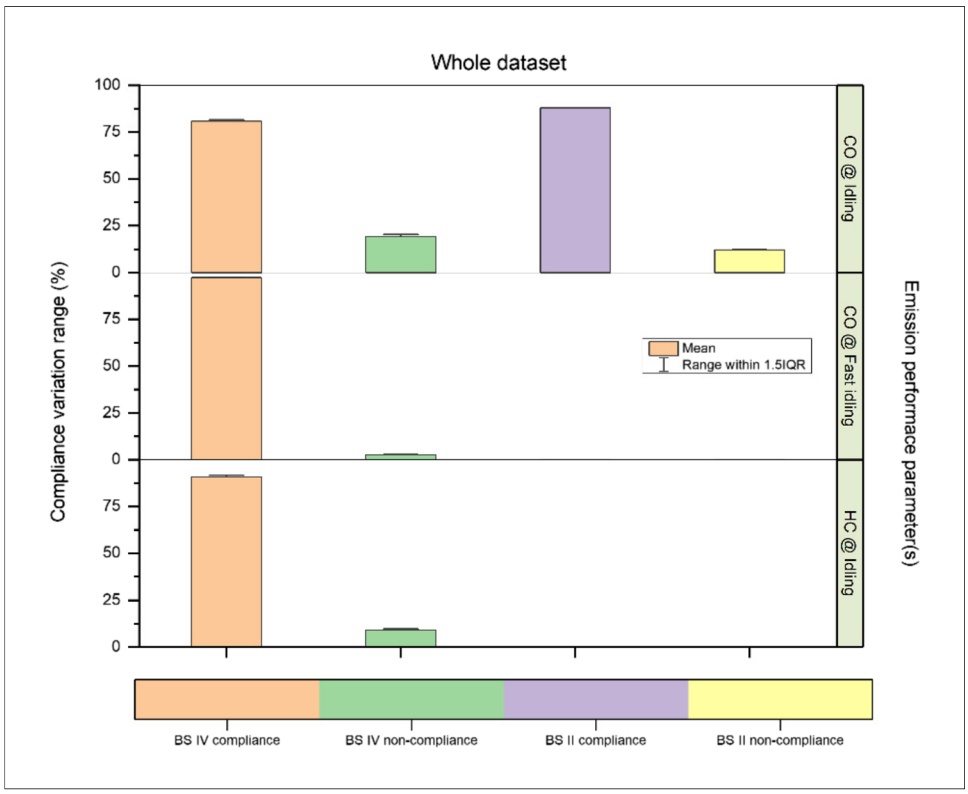


Fig. 1 Emission compliance levels for PDPCs (all datasets)

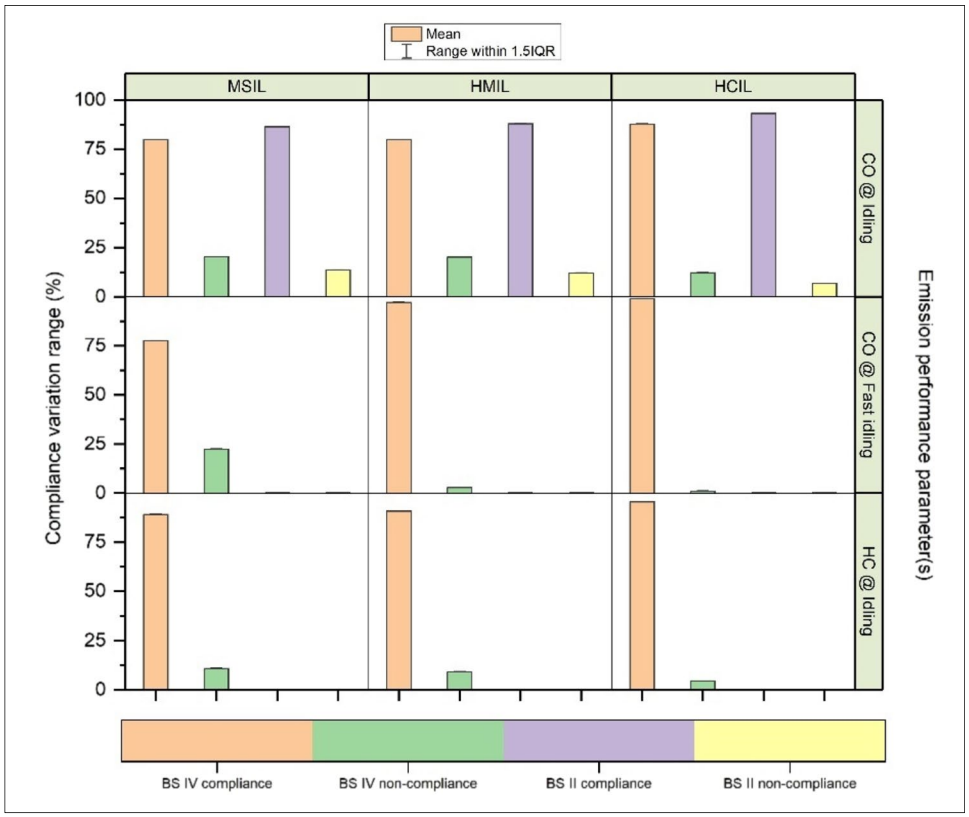


Fig. 2 Emission compliance levels for PDPCs (top 3 makes)

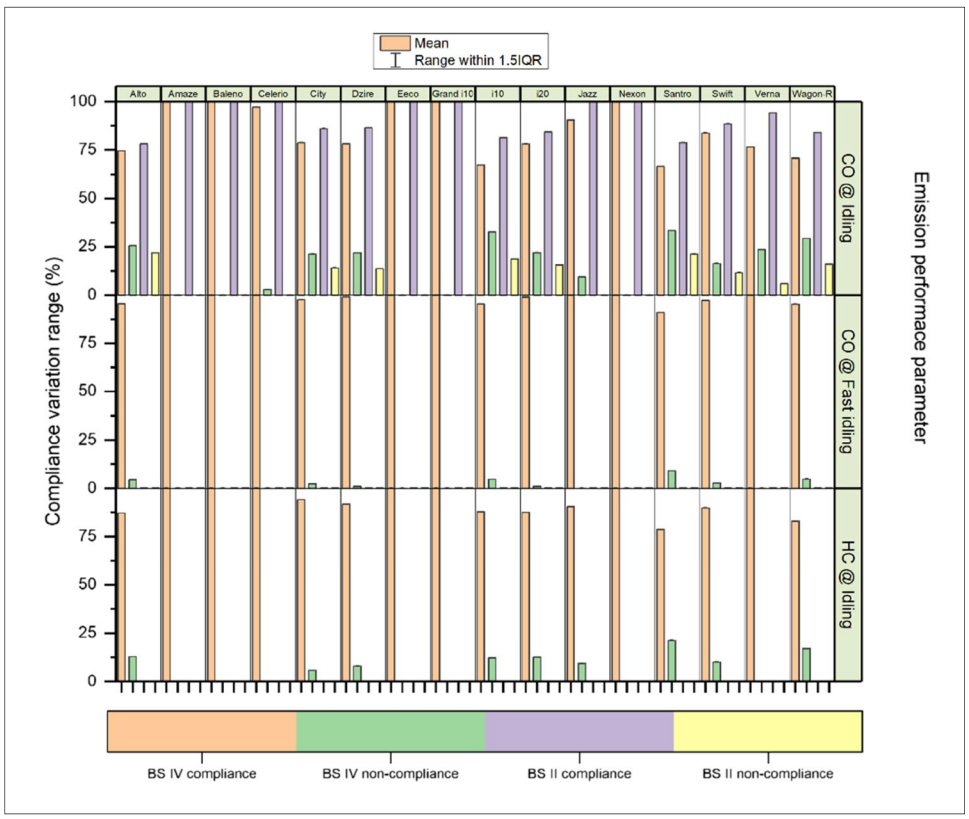


Fig. 3 Emission compliance levels for PDPCs (top 16 models)

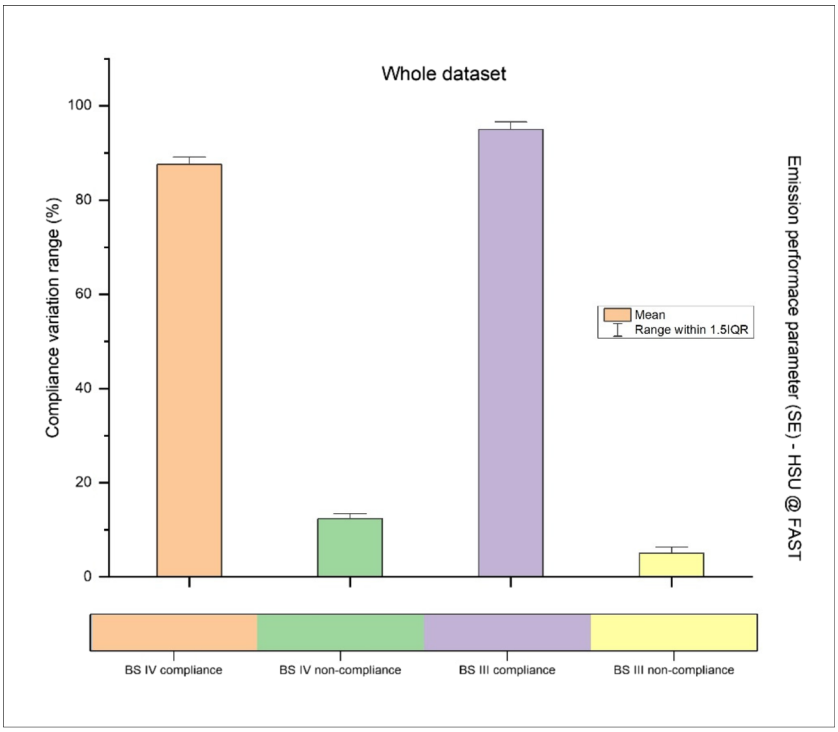


Fig. 4 Emission compliance levels for DDPCs (all datasets)

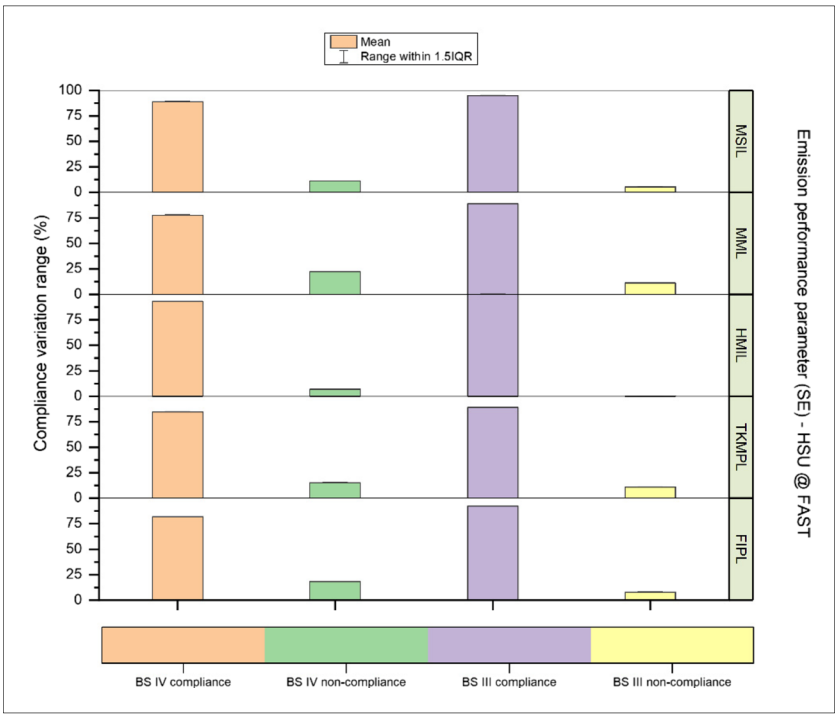


Fig. 5 Emission compliance levels for DDPCs (top 5 makes)

Again, all dataset combination discloses a very strong extent of conformance to both the emission norms (BS IV and BS III) for their smoke emission (reflected in terms of HSU) when subject to testing in the FAS test mode. Because the BS III emission standard is replaced by the relatively newer and more stringent BS IV (65 HSU in case of later, instead of 50 in the former's case), compliance of diesel cars toward BS III is substantially and understandably so greater in all three dataset situations. The dataset corresponding to the top 13 models of diesel cars also seems to do better than the remaining two datasets in terms of non-conformance to both the in-use

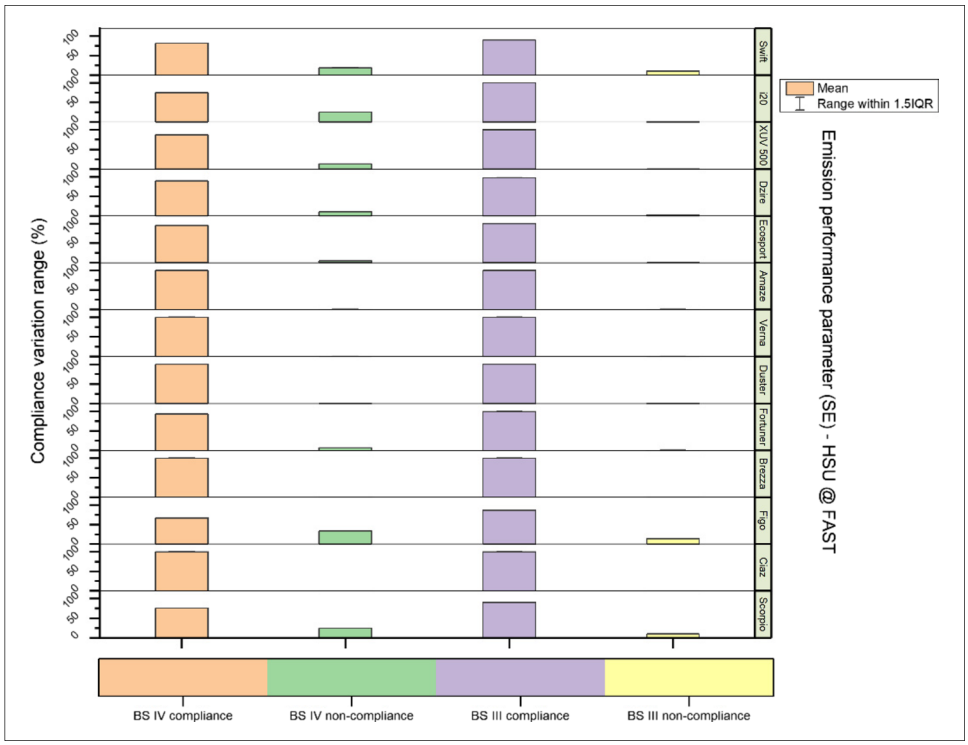


Fig. 6 Emission compliance levels for DDPCs (top 13 models)

emission standards, namely, BS III and BS IV, reflecting the lowest compliance values.

Make-wise speaking, the HMIL (Hyundai) displays the highest level of compliance toward both BS IV and BS II emission standards. MSIL (Maruti) and TKMPL (Toyota) follow closely with second-best compliance levels. MML (Mahindra) and FIPL (Ford) exhibit relatively superior non-compliance degrees, notably for BS IV. In terms of model dataset scenario, car models such as Duster, Verna, Brezza, Amaze, and Ciaz represent the greatest magnitude of conformance for both BS IV and BS II emission norms, recording virtually 100% compliance level. Figo emerges as the lowest performer having recorded the poorest compliance levels with i20, Scorpio, and Swift models following closely.

Analysis of vehicle condition (applicable emission norm and maintenance category)

The emission compliance level was analyzed for two vehicle-related aspects like emission norms during manufacturing and vehicle maintenance status. The boxplots depicting the scenario for emission norms used in the present study reveal that BS IV-compliant petrol-driven passenger cars emit the lowest range of CO in both idle and fast idle scenarios, while in the case of HC emission, the apparent high degree of outliers seen in emission norm parameter does not support this outcome. The effect of the maintenance category of vehicles has a more pronounced effect on CO and HC compliances with "very good" and "good" categories of vehicles always emitting within the BS IV emission norm (Figs. 7 and 8).

Benefitting from the wider range of BS II norms compared to BS IV, even better compliance is seen as only the "unsatisfactory" category vehicles emitted in violation of the BS II standard except HC. Make-wise compliance scenario for petrol cars is presented in Fig. 9, which reveals that HCIL exhibits the highest emission compliance levels or all emission test parameters (i.e., CO @ idling, fast idling, and HC @ idling for BS IV and BS II emission standards) followed by HMIL and MSIL makes. The maximum degree of non-compliance to both emission norms for all the testing parameters was depicted by MSIL make. The emission compliance levels analyzed for emission norms during manufacturing and vehicle maintenance parameters in the case of diesel-driven passenger cars are presented in Fig. 10. The boxplots show that BS IV-compliant diesel cars emit the lowest range of HSU emissions followed by BS II-compliant vehicles. Like petrol-driven cars, the maintenance category of vehicles is observed to have a more apparent effect on HSU emission, and except for the "unsatisfactory" category, all other vehicles possessing "very good" and "good" categories of vehicles always emit within the BS IV emission norm. Even "poor" category diesel cars were found to be compliant with BS IV norms, although their compliance is just on the threshold.

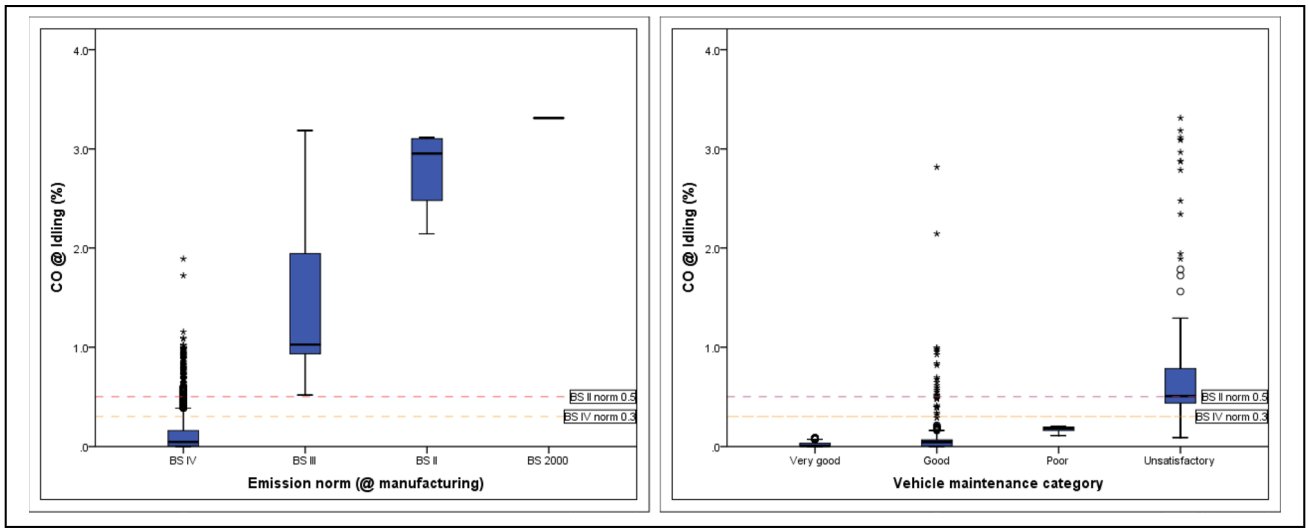


Fig. 7 Emission norm and maintenance category vs. CO emission (at idling)

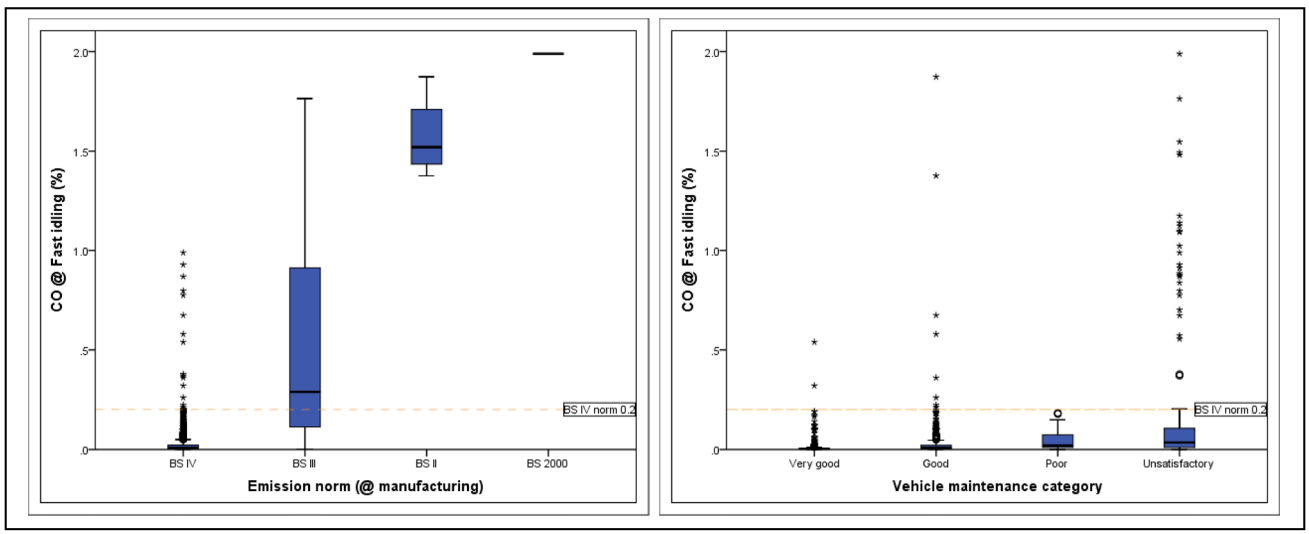


Fig. 8 Emission norm and maintenance category vs. CO emission (at fast idling)

Conclusion

The present study conducted a comprehensive assessment of actual compliance of both petrol and diesel-driven, privately owned passenger cars to the extant in-use emission standards as part of concurrent environmental policy and regulations. Such policy is prevalent in most South-East Asian countries, which revolves around certifying the emission compliance of the vehicle based on idle and high idle tailpipe CO and HC concentrations for petrol and SE values for diesel cars. Following the emission certification protocol prescribed under the CVMR, we investigated the 2040 number of cars plying on the roads of the NCT of Delhi, India. We carried out an in situ, exploratory data analysis keeping in aim the compliance level of vehicles to BS (Bharat Stage) emission norms for different datasets (entire top 3/top 5 makes and top 16/top 13 models). Additionally, we examined how two more car-related independent variables, namely, applicable in-use emission standard and car maintenance status affected the tailpipe CO, HC, and SE.

We reported all dataset scenarios of petrol cars to be highly compliant toward the BS IV norm (the latest and the most stringent at the time of the present tailpipe emission testing program) and even better toward BS II for

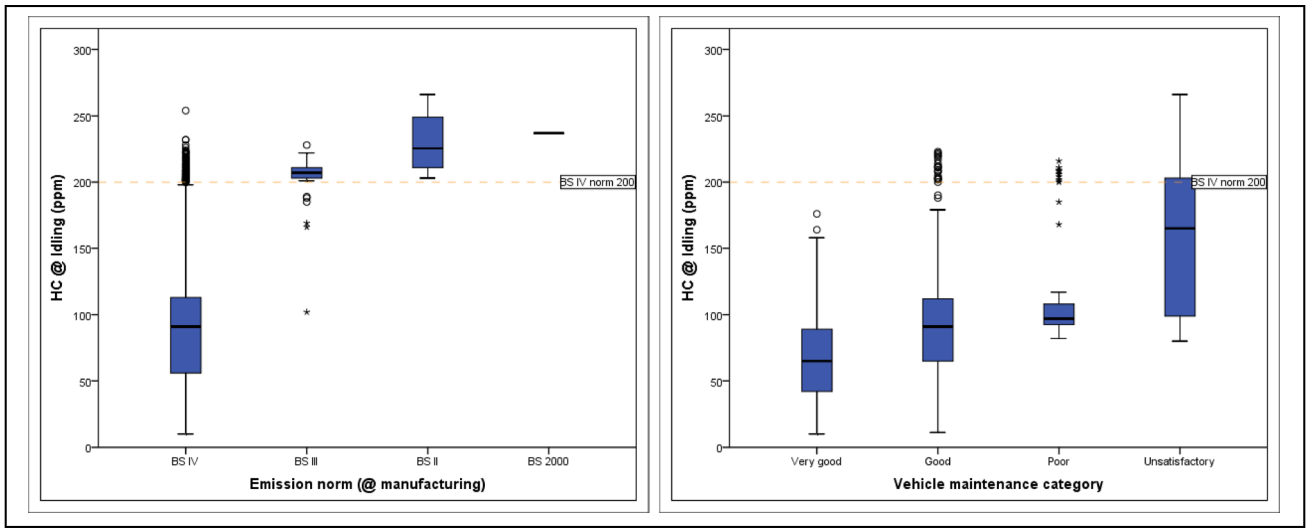


Fig. 9 Emission norm and maintenance category vs. HC emission (at idling)

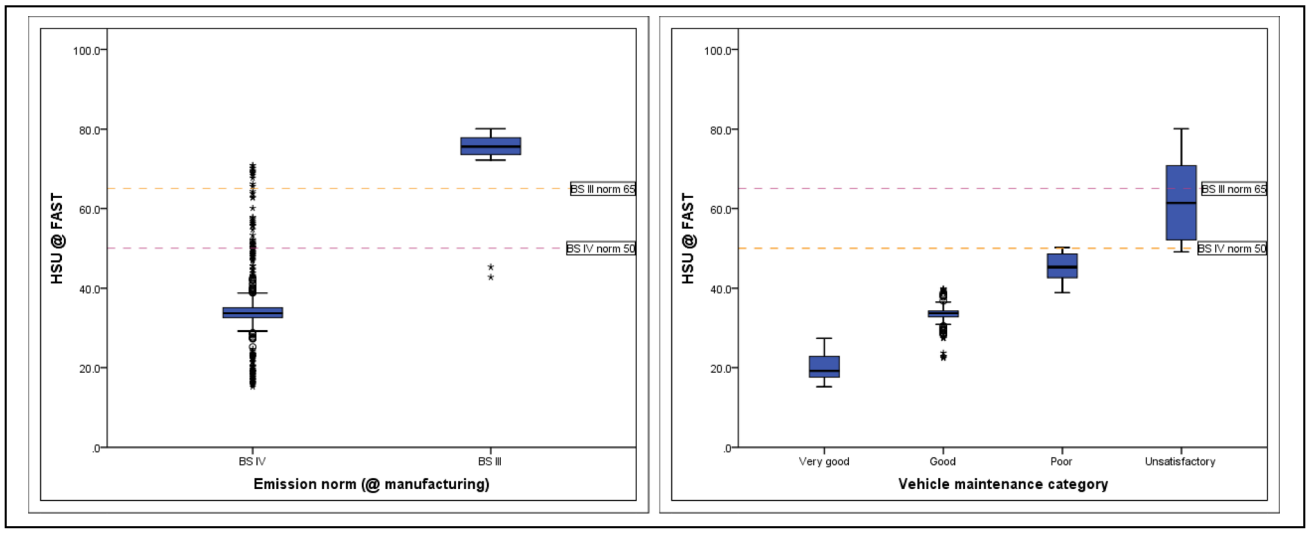


Fig. 10 Emission norm and maintenance category vs. HSU (at FAST)

CO and HC emissions (entire dataset and in case of top 3 makes 80–90%). The compliance to the HC norm is usually better than that to CO (85–90% against 75–80% in the case of both dataset scenarios) The reason is the wider range of the BS II (older) norm which lets a higher number of cars pass through. High idling CO compliance is found to be the best (> 90% for all datasets), except MSIL makes reporting non-compliance to the magnitude of 20%. Model-wise compliance status finds Amaze, Baleno, Eeco, Grand i10, and Nexon to be the most compliant models for various test parameters achieving almost 100% compliance toward BS IV norms.

For diesel-driven cars, conformity levels for BS III are evident to be superior to that for the BS IV standard, although not differing considerably (90% in the case of BS III while 80% in the case of BS IV). Make-wise MML and model-wise Figo are the least compliant models, whereas HMIL as make and Amaze and Verna are the most compliant ones. The study further reveals that the effect of the maintenance category of vehicles has a more pronounced effect on CO, HC, and SE compliances with "very good" and "good" categories of vehicles always emitting within the BS IV emission norm. Interestingly, the lenient idle testing recertification norms in developing countries have been

leading to a very high compliance rate (or very low degree of non-conformances), which might not be reflecting the actual emission performance status of the urban passenger cars' fleet. The study reiterates the inclusion of high idle testing (as on the interim basis) and mass emission (gm/km)-based assessment in the wholesome automobile emission evaluation and reduction policy, particularly concerning developing countries.

Further, we have highlighted the key parameters to be considered in the proposed emission certification policy in developing countries while also looking into lowering emission benchmarks (Fig. SM4). The study establishes an emission data bank which is particularly helpful for environmental regulation or policymakers in understanding the urban vehicular emission scenario. Driven by the finding that the condition of emission recertification instrumentation infrastructure in various countries remains poor, we suggest that interim measures be deployed so that driving conditions are mimicked to the extent possible, while the cars are tested for their emission compliance levels. The outcomes in relation to the present study will help policymakers and governments adopt suitable strategies before the cars become non-compliant with the corresponding in-use emission standards or help them remain in the longer status of compliance.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11356-024-32555-z.

Author contributions AP: methodology development, investigation, formal analysis, and original draft preparation and editing
GP: conceptualization, supervision, and critical inputs
RKM: supervision, critical inputs, and review and editing

Data availability Data will be made available on reasonable request.

Declarations

Ethics approval and consent to participate Not applicable

Consent for publication Not applicable

Competing interests The authors declare no competing interests.

References

Bandivadekar A, German J, Mock P (2015) Position brief policy solutions to reduce vehicle exhaust emissions under real-world driving conditions. https://theicct.org/publications/policy-solutions-reduce-vehicle-exhaust-emissionsunder-real-world-driving-conditions. Accessed 15 Aug 2020

Bella PR, Venkateswar R (2010) A review on diesel soot emission, its effect and control. Bull Chem React Eng Catal 5(2):69–86. https://doi.org/10.9767/bcrec.5.2.794.69-86

CPCB (2010) Status of the vehicular pollution control program in India. Central Pollution Control Board, Ministry of Environment, Forests and Climate Change, Delhi. https://cpcb.nic.in/archivereport.php. Accessed 11 May 2020

Dallman T (2020) Brazil PROCONVE L-7 and L-8 emission standards for light-duty vehicles. https://theicct.org/publications/brazil-proconve-l-7-and-l-8-emissionstandards-light-duty-vehicles. Accessed 20 Aug 2020

Dandapat S, Ghosh T, Shankar U, Maitra S, Maitra B (2020) A relook at the pollution certification of in-use vehicles in India and a way forward. Asian Transpo Stud 6:100020. https://doi.org/10.1016/j.eastsj.2020.100020

European Union (2023) Support for the in-service verification of CO_2 emissions of new light- and heavy-duty vehicles – final report. https://op.europa.eu/en/publication-detail/-/publication/0eb4bfc8-f080-11ed-a05c-01aa75ed71a1/language-en. Accessed 15 Dec 2023

Fernandes P, Ferreira E, Amorim P, Coelho MC (2022) Comparing different approaches for estimating tailpipe emissions in passenger cars. Transp Res Procedia 62:509–516. https://doi.org/10.1016/j.trpro.2022.02.063

Gulia S, Nagendra SS, Barnes J, Khare M (2018) Urban local air quality management framework for non-attainment areas in Indian cities. Sci Total Environ 619:1308–1318. https://doi.org/10.1016/j.scitotenv.2017.11.123

Gulia S, Tiwari R, Mendiratta S, Kaur S, Goyal SK, Kumar R (2020) Review of scientific technology-based solutions for vehicular pollution control. Clean Techn Environ Policy 22:1955–1966. https://doi.org/10.1007/s10098-020-01952-6

He H, Jin-Beijing L, Francisco S, Albu S, Freed C, Maxwell R, Oge M, Simon K, Urkov J, Walsh MP, Cackette T, Ding Y, Tang D, German J, Kamakate F, Muncrief R, Yang Z (2017) A historical review of the U.S. vehicle emission compliance program and emission recall cases. https://theicct.org/publications/historical-review-us-vehicle-emission-compliance-program-and-emission-recall-cases. Accessed 8 Oct 2020

Karthikeyan S, Prathima A (2016) Environmental effect of CeO_2 nano-additive on biodiesel. Energy Sources A: Recovery Util Environ Eff 38(24):3673–3679. https://doi.org/10.1080/15567036.2016.1177624

Kumar P, Gulia S, Harrison RM, Khare M (2017) The influence of odd–even car trial on fine and coarse particles in Delhi. Environ Pollut 225:20–30. https://doi.org/10.1016/j.envpol.2017.03.017

Lyu M, Bao X, Zhu R, Matthews R (2020) State-of-the-art outlook for light-duty vehicle emission control standards and technologies in China. Clean Techn Environ Policy 22(4):757–771. https://doi.org/10.1007/s10098-020-01834-x

Maxwell R, Hannon J (2017) U.S. EPA recall program for light-duty vehicles and light-duty trucks. https://theicct.org/publications/us-epa-recall-program-lightduty-vehicles-and-light-duty-trucks. Accessed 5 Jan 2021

Mock P, German J (2015) The future of vehicle emissions testing and compliance how to align regulatory requirements, customer expectations, and environmental performance in the European Union. https://theicct.org/sites/default/files/publications/ICCT_future-vehicletesting_20151123.pdf. Accessed 5 Jan 2021

MoRTH (2020) Central motor vehicle rules. Ministry of Road Transport and Highways, Government of India. https://morth.nic.in/sites/default/files/notifications_document/G.S.R498%28E%29.pdf. Accessed 30 Jun 2022.

Nesamani KS (2010) Estimation of automobile emissions and control strategies in India. Sci Total Environ 408:1800–1811. https://doi.org/10.1016/j.scitotenv.2010.01.026

Ogunseye OO, Ana GREE, Uhiara DC, Shendell DG (2018) Carboxyhemoglobin levels among traders exposed to vehicular emissions in three motor parks in Ibadan, Nigeria. J Environ Public Health 1–8. https://doi.org/10.1155/2018/9174868

OICA (2019) Global sales statistics of passenger cars. https://www.oica.net/category/sales-statistics/. Accessed 30 Jul 2020

Pandey A, Mishra RK, Pandey G (2022a) Investigating exhaust emission from in-use passenger cars: an exploratory analysis and policy outlook. J Environ Eng (New York) 148(7). https://doi.org/10.1061/(ASCE)EE.1943-7870.0002015

Pandey A, Pandey G, Mishra RK (2022b) Applying the indexing system for assessment of effectiveness of the exhaust emission compliance certification process for passenger cars. Proc Math Phys Eng Sci 478(2226). https://doi.org/10.1098/rspa.2022.0315

Raparthi N, Phuleria HC (2022) Assessing Mumbai's in-use vehicular characteristics, current emissions, and future projections under various policy interventions. J Clean Prod:134145. https://doi.org/10.1016/j.jclepro.2022.134145

Ribeiro CB, Rodella FHC, Hoinaski L (2021) Regulating light-duty vehicle emissions: an overview of US, EU, China and Brazil programs and its effect on air quality. Clean Techn Environ Policy. https://doi.org/10.1007/s10098-021-02238-1

Rice MB, Ljungman PL, Wilker EH, Dorans KS, Gold DR, Schwartz J, Koutrakis P, Washko GR, O'Connor GT, Mittleman MA (2015) Long-term exposure to traffic emissions and fine particulate matter and lung function decline in the Framingham Heart Study. Am J Respir Crit Care Med 191:656–664. https://doi.org/10.1164/rccm.201410-1875oc

Rodríguez F, Bernard Y, Dornoff J, Mock P (2019) Recommendations for post-euro 6 standards for light-duty vehicles in the European Union. https://theicct.org/publications/recommendations-post-euro-6-eu. Accessed 11 Feb 2021

Sharma R, Kumar R, Singh PK, Raboaca MS, Felseghi RA (2020) A systematic study on the analysis of the emission of CO, CO_2 and HC for four-wheelers and its impact on the sustainable ecosystem. Sustainability 12(17):6707. https://doi.org/10.3390/su12176707

Sharma S, Marechal F (2019) Carbon dioxide capture from internal combustion engine exhaust using temperature swing adsorption. Front Energy Res 7:143. https://doi.org/10.3389/fenrg.2019.00143

SIAM (2018) In-use vehicle emission norms. Society of Indian Automobile Manufacturers. https://www.siam.in/technical-regulation.aspx?mpgid=31&pgidtrail=35. Accessed 12 Apr 2022

Team BHP (2022) Passenger car sales data in India. https://www.team-bhp.com/forum/indian-car-scene/231790-2020-report-card-annual-indian-car-sales-analysis.html. Accessed 10 Dec 2022

Ventura LM, Pinto FO, Gioda A, D'Agosto MA (2020) Inspection and maintenance programs for in-service vehicles: an important air pollution control tool. Sustain Cities Soc 53:101956. https://doi.org/10.1016/j.scs.2019.101956

Wang H, Fu L, Zhou Y, Du X, Ge W (2010) Trends in vehicular emissions in China's mega cities from 1995 to 2005. Environ Pollut 158:394–400. https://doi.org/10.1016/j.envpol.2009.09.002

Wang S, Xing J, Zhao B, Jang C, Hao J (2014) Effectiveness of national air pollution control policies on the air quality in metropolitan areas of China. J Environ Sci (China) 26(1):13–22. https://doi.org/10.1016/s1001-0742(13)60381-2

WHO (2019) The global health observatory. https://www.who.int/data/gho/data/themes/theme-details/GHO/air-pollution. Accessed 15 Sept 2020

Wong YK, Huang XHH, Cheng YY, Louie PKK, Yu ALC, Tang AWY, Chan DHL, Yu JZ (2019) Estimating contributions of vehicular emissions to $PM_{2.5}$ in a roadside environment: a multiple approach study. Sci Total Environ 672:776–788. https://doi.org/10.1016/j.scitotenv.2019.03.463

Zhong C, Shen S, Li J, Yang Z (2022) The research of the inspection and maintenance program and its implementation. E3S Web Conf 360:01014. https://doi.org/10.1051/e3sconf/202236001014

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.

RSC Advances

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *RSC Adv.*, 2023, **13**, 33675

Energy transfer mechanisms and color-tunable luminescence of $Tm^{3+}/Tb^{3+}/Eu^{3+}$ co-doped $Sr_4Nb_2O_9$ phosphors for high-quality white light-emitting diodes

Ravina,[a] Kanishk Poria,[b] Mukesh K. Sahu,[c] A. Kumar,[a] Anu,[d] Sajjan Dahiya,[e] Nisha Deopa *[a] and A. S. Rao[d]

This paper investigates the synthesis and luminescence characteristics of $Tm^{3+}/Tb^{3+}/Eu^{3+}$ co-doped $Sr_4Nb_2O_9$ (SNB) phosphors as potential candidates for white light-emitting diodes (WLEDs). The study explores the energy transfer mechanisms and color-tunable characteristics of these phosphors. The SNB phosphors were prepared using a solid-state reaction method, and their structural and morphological properties were characterized using X-ray diffraction (XRD), scanning electron microscopy (SEM), energy-dispersive X-ray spectroscopy (EDS), and Fourier-transform infrared (FT-IR) spectroscopy. The diffuse reflectance, photoluminescence (PL) and time resolved photoluminescence (TRPL) properties were investigated, revealing efficient energy transfer processes from Tm^{3+} to Tb^{3+} and Eu^{3+} ions. The energy transfer mechanisms were determined through critical distance calculations and analysis of multipolar interactions. The co-doped phosphors exhibited tunable emission colors ranging from blue to white light, with controllable correlated color temperatures (CCTs) and high color rendering indices (CRIs). The CIE chromaticity coordinates were optimized to approach neutral white light. The PL intensity is maintained at 81.19% at 150 °C of that of room temperature which showcases the remarkable thermal stability of the as-prepared phosphors. The results highlight the potential of $Tm^{3+}/Tb^{3+}/Eu^{3+}$ co-doped SNB phosphors for generating high-quality, color-tunable white light for advanced lighting applications.

Received 14th August 2023
Accepted 8th November 2023

DOI: 10.1039/d3ra05519a

rsc.li/rsc-advances

Open Access Article. Published on 16 November 2023. Downloaded on 11/18/2023 11:30:10 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

1 Introduction

With technological advancements, especially the widespread utilization of rare earth ion-doped phosphors, white light-emitting diodes (WLEDs) have progressively emerged as the optimal choice for next-generation lighting solutions. This preference arises from their inherent advantages, including cost-effectiveness, high efficiency, and environmental friendliness.[1–3] Currently, white light can be achieved through two approaches. The first involves integrating a blue InGaN chip with YAG:Ce^{3+} yellow phosphors. However, this approach lacks the red segment of the spectrum, leading to a diminished color rendering index (CRI) and an elevated correlated color temperature (CCT). This deviation from the natural white light region poses a challenge to achieving high-quality white light illumination. Another approach involves using a near-ultraviolet chip combined with three-color phosphors. This method is capable of meeting the requirements for a high color rendering index (CRI) and achieving a suitable correlated color temperature (CCT).[4] By employing this approach, it becomes possible to produce white light that closely resembles natural lighting conditions. Therefore, the development of color-tunable phosphors on a single-phase substrate is a highly significant research direction within the field of white light-emitting diodes (WLEDs). This approach aims to create phosphors that can be adjusted to produce a wide range of colors, enabling versatile and customizable lighting solutions.

Rare-earth ions exhibit a remarkably diverse electronic structure, rendering them capable of emitting photoluminescence across a wide range of wavelengths, including visible and infrared light. Particularly in the visible light region, their emission ability and vivid colors stand out, making rare-earth ions highly desirable for various applications, such as lighting and display technologies.[5] The fusion of three primary colors—red, green, and blue—is well-known to yield white light.

[a]*Department of Physics, Chaudhary Ranbir Singh University, Rohtak Bypass Road, Jind-126102, Haryana, India. E-mail: nispectro999@gmail.com; Tel: +918860551723*
[b]*Department of Physics, Panjab University, Chandigarh-160014, India*
[c]*School of Computing, Graphic Era Hill University, Haldwani, Uttarakhand-263136, India*
[d]*Department of Applied Physics, Delhi Technological University, Bawana Road, New Delhi-110 042, India*
[e]*Department of Physics, Maharshi Dayanand University, Rohtak-124001, Haryana, India*

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 16 November 2023. Downloaded on 11/18/2023 11:30:10 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

In this context, different phosphors utilize Tm^{3+}, Eu^{3+}, and Tb^{3+} ions as blue, red, and green emitters, respectively. These ions are chosen for their abundant electron transitions, contributing to the emission properties of the phosphors. For instance, Tm^{3+} doping in inorganic compounds often yields excellent blue-emitting performance due to the blue transition from $^1D_2 \rightarrow {}^3F_4$.[6,7] Eu^{3+} ions emit in the orange-red region of the spectrum through the $^5D_0 \rightarrow {}^7F_J$ ($J = 0$–4) transitions,[8,9] while Tb^{3+} ions are commonly used as blue-green emitters with the $^5D_4 \rightarrow {}^7F_J$ ($J = 6$–3) transitions.[10] Co-doping Tm^{3+}, Eu^{3+}, and Tb^{3+} ions into a single host material allows for the potential generation of a single-phased phosphor emitting white light.[11] This approach permits the generation of white light by combining emissions from these different ions. Furthermore, the selection of an appropriate host material is vital for synthesizing phosphors with high performance. The host material plays a crucial role by providing suitable crystal fields for the activator ions. The crystal structure and properties of the host material significantly influence the luminescent properties and efficiency of the phosphor. Hence, careful consideration and optimization of the host material are essential to achieve phosphors with desirable performance characteristics.

The extensive study of niobate group materials is driven by their remarkable electro-optic, dielectric, ferroelectric, pyroelectric, and piezoelectric properties.[12,13] Niobate serves as a highly favorable host matrix for doping rare-earth ions due to its exceptional luminescent properties, excellent chemical stability, and good thermal stability. Concerning characteristics related to light emission, the optical absorption observed in niobates is attributed to charge transfer occurring within the NbO_6 complex.[14] This absorption behavior significantly influences their luminescent properties. While certain pure niobates may not exhibit luminescence, substantial research has been conducted on rare-earth-doped niobates due to their potential as light-emitting materials. The incorporation of rare-earth ions into niobate structures allows for the creation of luminescent centers, expanding the range of applications in areas such as lighting, displays, and optoelectronics.

2 Experimental procedure and instrumentations

Through the solid-state reaction method, we synthesized samples of strontium niobate phosphor $Sr_4Nb_2O_9$ (SNB) doped with Tm^{3+}, Tb^{3+}, and Eu^{3+}. The precursor materials $SrCO_3$ (98%), Nb_2O_5 (99.9%), Tm_2O_3 (99.9%), Tb_4O_7 (99.9%), Eu_2O_3 (99.99%) were of analytical grade. Following their precise stoichiometric ratio, we weighed and thoroughly mixed the precursors using an agate mortar until achieving uniformity. These homogeneous mixtures were then poured in an alumina crucible and sintered at 1350 °C for 7 hours, followed by gradual cooling to ambient temperature. We labeled the resulting phosphors as follows: SNB:xTm^{3+} ($x = 0.01$, 0.03, 0.05, 0.07, 0.10 mol), SNB:0.03Tm^{3+}, yTb^{3+} ($y = 0.01$, 0.02, 0.03, 0.04, 0.05, 0.06, 0.08, 0.10 mol) denoted as TT1-TT10, SNB:0.03Tm^{3+}, 0.06Tb^{3+}, zEu^{3+} ($z = 0.01$, 0.03, 0.05 and 0.07 mol) labeled as TTE1-TTE7. Finally, we finely powdered the prepared samples for subsequent characterization.

Structural analysis of the samples was conducted using X-ray diffraction (XRD). A Bruker D8 Advance diffractometer equipped with a nickel filter and a Cu Kα radiation source (wavelength $\lambda = 1.5406$ Å) was utilized within the 2θ range of 20° to 80°. For examination of sample morphology and elemental composition, scanning electron microscopy (SEM) and energy-dispersive X-ray spectroscopy (EDS) were performed using a Zeiss Gemini 300 instrument. Additionally, Fourier-transform infrared (FT-IR) spectra of the as-synthesized samples were recorded using a Nicolet IS50 FT-IR apparatus. UV spectrum analysis was conducted using a Jasco V-770 spectrophotometer with light sources of deuterium and halogen and a resolution of 1.0 nm. For validation of PL excitation and emission spectra, a Jasco 8300 FP spectrofluorophotometer employing 300 W xenon lamp acts as excitation source and a resolution of 1.0 nm was used. PL decay measurements were performed using the Hitachi F 7000 spectrofluorometer in conjunction with a 450 watts microsecond xenon flash lamp. The time-resolved photoluminescence (TDPL) was performed using the FLAME-S-XR1-ES Ocean Optics spectrometer along with a sample holder along with a heating assembly.

3 Results and discussion

3.1. Study of XRD pattern

Fig. 1 presents the XRD patterns of SNB phosphors doped with 3 mol% Tm^{3+}, 3 mol% Tm^{3+} and 6 mol% Tb^{3+}. The diffraction patterns demonstrate consistency with standard data (Card No. 048-0558), confirming the successful incorporation of Tm^{3+}, Tb^{3+} and Eu^{3+}ions into the host lattice. The XRD patterns confirm that the as-prepared phosphor possesses a crystalline cubic structure within the *Fm*3*m* space group, showing no structural alterations.[15] The impurities at 28° and 50° in the XRD is because of the difference between both ionic radius and charge numbers between the Ln^{3+} ions and Sr^{2+} ions. The average crystallite sizes (D) were calculated using Debye–Scherer equation (eqn (1)) from the literature,[16]

$$D = \frac{K\lambda}{\beta \cos \theta} \quad (1)$$

resulting in values of 41.64 nm for SNB:3 mol% Tm^{3+}, 40.61 nm for SNB:3 mol% Tm^{3+} and 6 mol% Tb^{3+}, and 41.07 nm for SNB:3 mol% Tm^{3+}, 6 mol% Tb^{3+}, and 5 mol% Eu^{3+}.

Fig. 2 demonstrates the XRD patterns of the SNB:3Tm^{3+}/6Tb^{3+}/zEu^{3+} (z = 1, 3, 5 and 7 mol%) samples which are consistent with the standard data. The average crystallite size were evaluated and found to be 42.26, 41.88, 41.07 and 42.12 nm, respectively.

3.2. SEM and EDS studies

SEM imaging, SEM-EDS mapping, and EDS spectrum analysis were employed to investigate the morphology and elemental composition of the SNB sample doped with 3 mol% Tm^{3+} and 1 mol% Tb^{3+}, as illustrated in Fig. 3. In Fig. 3(a), the SEM image

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 16 November 2023. Downloaded on 11/18/2023 11:30:10 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

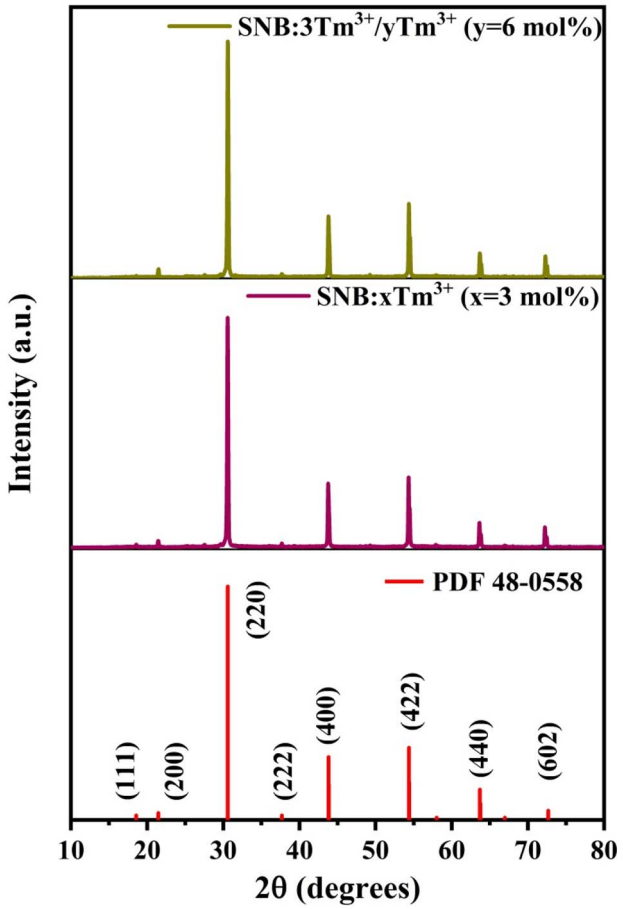


Fig. 1 XRD patterns of SNB:xTm^{3+} (x = 3 mol%), SNB:3Tm^{3+}/yTb^{3+} (y = 6 mol%) samples and compared with standard data.

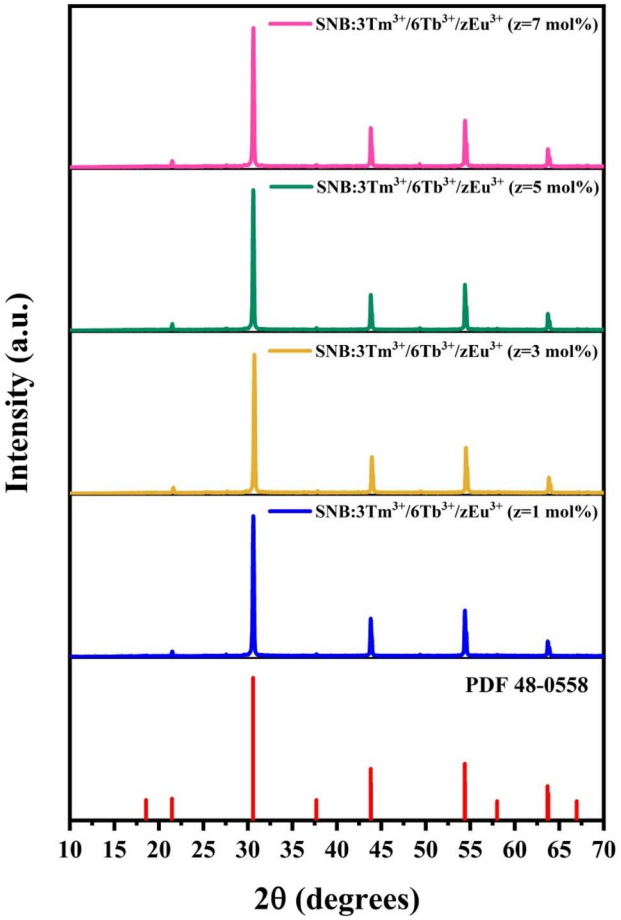


Fig. 2 XRD patterns of SNB:3Tm^{3+}/6Tb^{3+}/zEu^{3+} (z = 1, 3, 5 and 7 mol%) samples and compared with standard data.

unveils the presence of regular and smooth micron-sized particles within the SNB sample. Elemental mapping results in Fig. 3(b–f) exhibit a uniform distribution of Sr, Nb, O, Tm, and Tb elements across the grains. The EDS spectrum depicted in Fig. 3(g) displays distinct peaks corresponding to Nb, O, and Sr, along with smaller peaks representing Tm and Tb. No indications of impurity elemental peaks are evident, indicating the successful integration of Tm^{3+} and Tb^{3+} ions into the SNB host matrix. The inset within Fig. 3(g) illustrates the elemental compositions.

3.3. FTIR studies

The structural components and functional groups present in the as-synthesized SNB:xTm^{3+} (x = 1, 3, 5, 7, and 10 mol%) phosphor were analyzed by recording FT-IR spectra in the wavenumber range of 4000–400 cm^{-1}. Fig. 4 illustrates the FT-IR spectra of the synthesized phosphor. FT-IR spectroscopy was employed in the initial stage of material analysis to ensure the quality of the product. The shift in peak positions indicated the presence of contaminants. The spectra exhibited analogous positions of shoulder, as shown in Fig. 4. The FT-IR spectra revealed the presence of nine peaks at approximately 418, 435, 525, 651, 861, 1442, 2885, 2948, and 2989 cm^{-1}. These peaks were assigned based on previous reported papers. The peak observed near 418 and 525 cm^{-1} corresponds to the Sr–O stretching modes.[17] The peaks around 435 and 861 cm^{-1} are attributed to the coupling between Nb–O stretching mode.[18] The peak at approximately 651 cm^{-1} is due to Tm–O stretching vibrations.[19] The peaks observed in the range of 1442 and 2989 cm^{-1} are associated with water molecules.[20] The disturbances in the form of peaks between 1000 to 3000 cm^{-1} may indicate the presence of residual organic compounds.[21] Furthermore, the absence of significant evidence of carbonates or carboxyl groups in the FT-IR spectra confirms the synthesis of phosphor materials from their precursor carbonates.

3.4. Diffuse reflectance study

Diffuse reflectance spectra were collected for the SNB samples doped with 3 mol% Tm^{3+} and SNB:3 mol% Tm^{3+}, 6 mol% Tb^{3+} over the wavelength range of 200–1800 nm, spanning the UV-visible and near-IR regions at room temperature, as depicted in Fig. 5.

In the phosphors doped solely with Tm^{3+}, the absorption bands located at approximately 684, 788, 1216, and 1644 nm

Open Access Article. Published on 16 November 2023. Downloaded on 11/18/2023 11:30:10 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

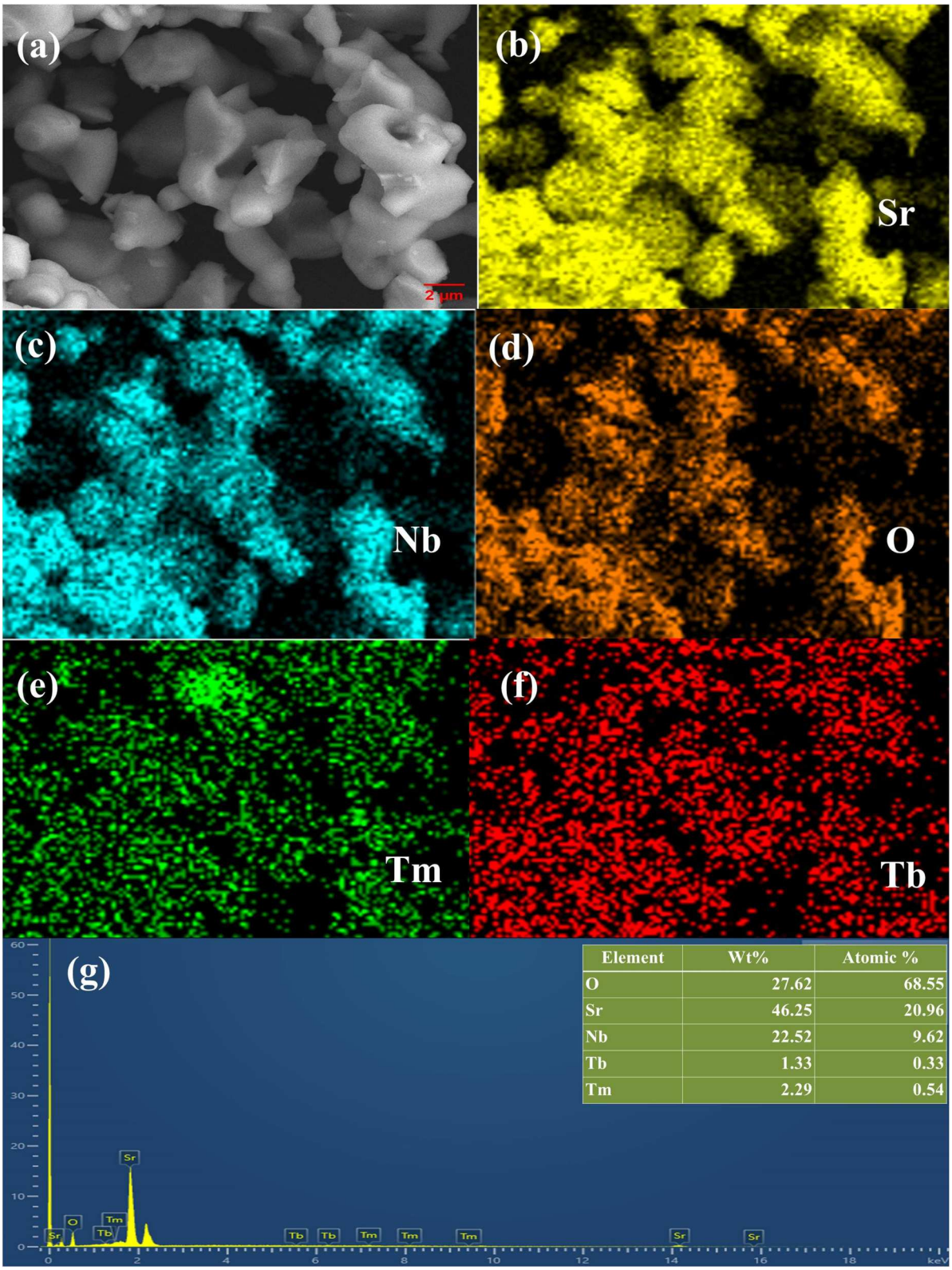


Fig. 3 (a) SEM images. (b–f) SEM-EDS mapping. (g) EDS spectrum of SNB:3 mol% Tm^{3+}, 6 mol% Tb^{3+} phosphor [inset shows the elements weight and atomic percent].

correspond to transitions from the ground level $^3H_6$ to $^3F_3$, $^3H_4$, $^3H_5$ and $^3F_4$ levels, respectively.[22] In phosphors co-doped with both Tm^{3+} and Tb^{3+}, a captivating interplay of energy level transitions is observable. Tm^{3+} ions exhibit transitions at 684, 788, 1216, and 1644 nm, while Tb^{3+} ions exhibits the $^7F_6 \rightarrow {}^5D_3$ energy level transition at 386 nm.[23]

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 16 November 2023. Downloaded on 11/18/2023 11:30:10 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

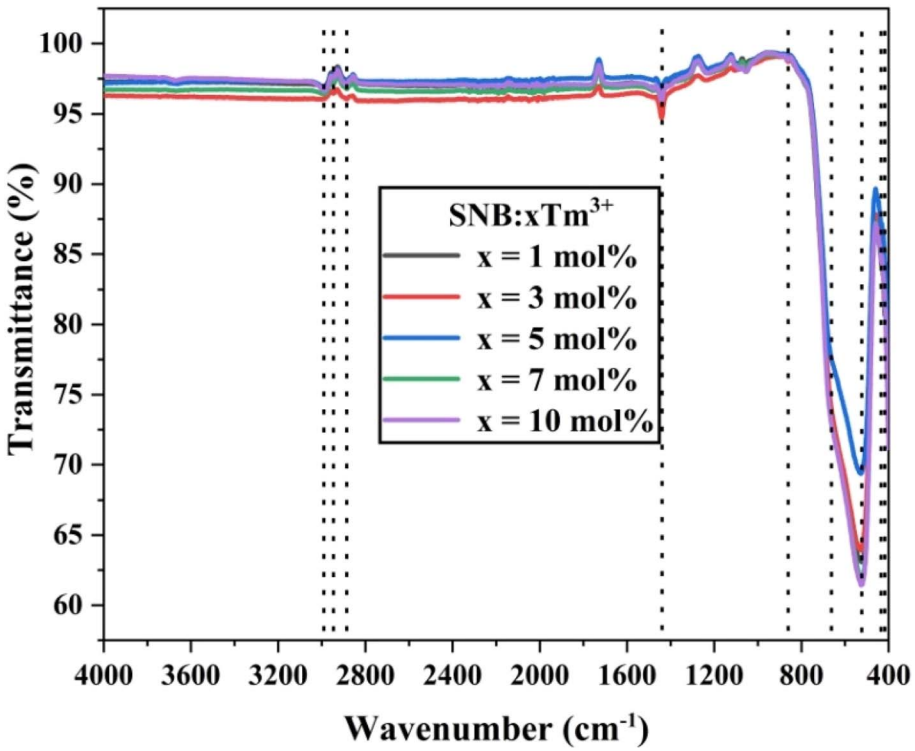


Fig. 4 FT-IR spectra of SNB:xTm^{3+} samples.

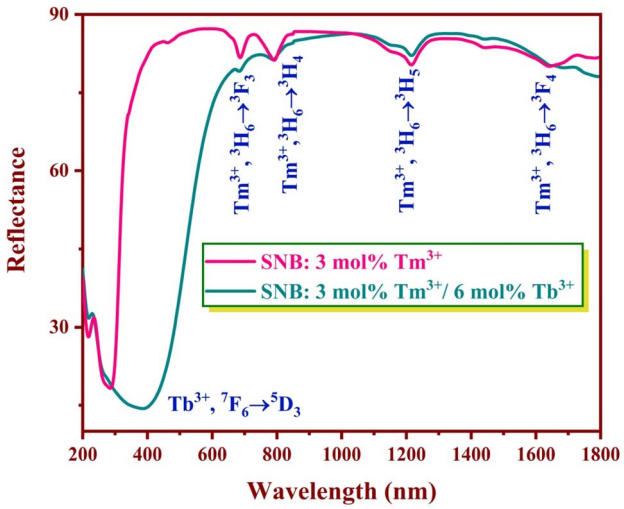


Fig. 5 Diffuse reflectance spectra of SNB:3 mol% Tm^{3+} and SNB:3 mol% Tm^{3+}, 6 mol% Tb^{3+} phosphors.

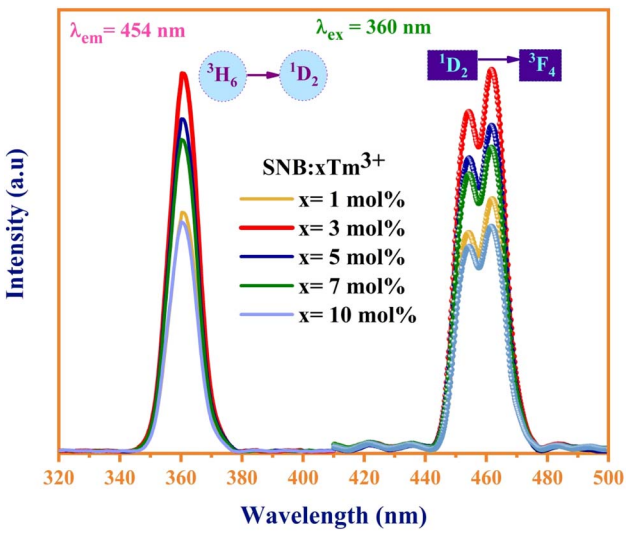


Fig. 6 PLE and PL spectra of SNB:xTm^{3+} phosphors.

3.5. Photoluminescence properties

Fig. 6 displays the photoluminescence excitation and emission spectra of the synthesized SNB:xTm^{3+} (x = 1, 3, 5, 7, 10 mol%) phosphors. The dominant excitation peak at 360 nm can be attributed to the distinctive $^3H_6 \rightarrow {}^1D_2$ transitions of Tm^{3+} ions, effectively stimulated by near-ultraviolet (NUV) light. In the emission spectra, a broad peak ranging from 440 to 480 nm under the 360 nm excitation for SNB:xTm^{3+} (x = 1, 3, 5, 7, 10) phosphors is associated with the $^1D_2 \rightarrow {}^3F_4$ transition.[24]

Interestingly, the peak intensity of the samples initially increases and then diminishes as the Tm^{3+} doping concentration rises. This intensity attains its peak at x = 3 mol%, which is attributed to concentration quenching, as depicted in Fig. 6. This phenomenon arises from the cross relaxation mechanism due to variations in the spacing between luminescence centers. At lower concentrations of Tm^{3+} ions, the spacing between them is sufficiently small to effectively neglect fluorescence quenching.[25] However, with an increasing concentration of Tm^{3+} ions,

Open Access Article. Published on 16 November 2023. Downloaded on 11/18/2023 11:30:10 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

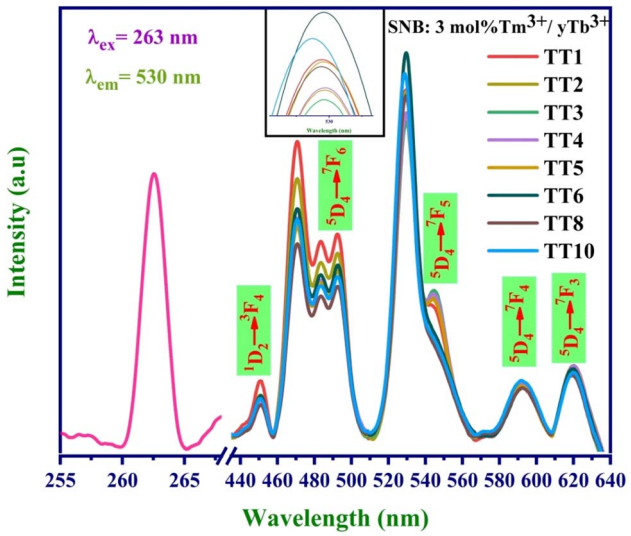


Fig. 7 PLE and PL spectra of SNB:3 mol% Tm^{3+}/yTb^{3+} phosphors.

the gap between the ions diminishes, falling below the critical transfer distance. This results in a higher utilization of Tm^{3+} ions at the $^1D_2$ level, consequently leading to a decrease in luminous efficiency.

3.6. Energy transfer from Tm^{3+} to Tb^{3+}

To explore the energy transfer mechanism from Tm^{3+} to Tb^{3+} ions, we prepared SNB phosphors doped with 3 mol% Tm^{3+} and varying concentrations of Tb^{3+} (denoted as y mol%, where $y = 1$, 2, 3, 4, 5, 6, 8, and 10 mol%). Fig. 7 illustrates the emission spectra under 263 nm UV light excitation, and a spectral comparison reveals that the emission peak positions remain nearly consistent across the 8 samples, appearing at approximately 450, 492, 530, 591, and 620 nm.[26]

Analysis of the emission spectra indicates that as the concentration of Tb^{3+} ions increases, the luminescent intensity of the distinct blue emission peak at 450 nm decreases. Meanwhile, the emission peaks attributed to Eu^{3+} ions at 530 nm initially experience an increase, reaching a maximum for Tm-3 mol% and Tb-6 mol%, followed by a gradual decrease. This observation suggests the occurrence of a concentration quenching phenomenon, signifying the energy transfer from Tm^{3+} to Tb^{3+} ions.

The efficiency of energy transfer from Tm^{3+} to Tb^{3+} ions can be quantified using the formula:[27]

$$\eta = 1 - \frac{I_T}{I_T + I_D} \quad (2)$$

In this equation, I_T and I_D denote the photoluminescence intensities of the emission peaks at 450 and 530 nm, respectively.

The calculated efficiency values for the TT1, TT2, TT3, TT4, TT5, TT6, TT8, and TT10 phosphors are 86.1%, 88.5%, 89.7%, 90.2%, 90%, 91.3%, 90.3%, and 90.4%, respectively. These results underscore a significant improvement in energy transfer efficiency with increasing concentration, reaching its peak value in the TT6 phosphor sample. The higher concentration of Tb^{3+} ions surrounding the Tm^{3+} ions promotes a greater likelihood of cross relaxation between the transitions of Tm^{3+} ($^1D_2 \rightarrow {}^3F_4$) and Tb^{3+} ($^5D_4 \rightarrow {}^7F_5$).[28] This phenomenon elucidates the concentration quenching and the reduction in emission intensity at 450 nm. The quantum yield for the TT6 phosphor was evaluated using the equations mentioned in literature and comes out to be 41% which is higher than the other reported works.[29,30]

Fig. 8 graphically portrays the energy transfer efficiency between Tm^{3+} and Tb^{3+} ions, alongside the emission intensities corresponding to the $^1D_2 \rightarrow {}^3F_4$ transition of Tm^{3+} and the $^5D_4$

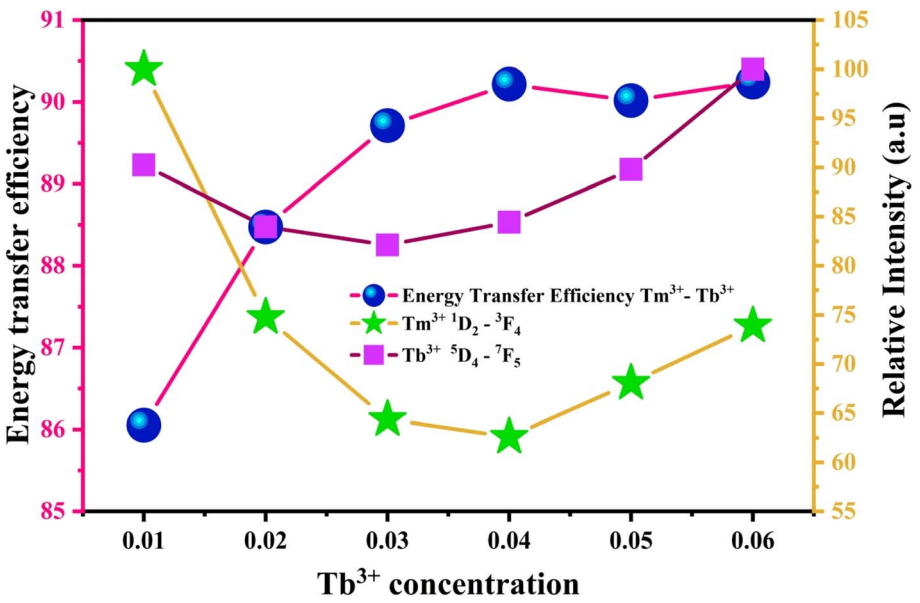


Fig. 8 Energy transfer efficiency of $Tm^{3+} \rightarrow Tb^{3+}$ and the luminous intensity curve of the monitoring wavelength were 450 nm (Tm^{3+}) and 530 nm (Tb^{3+}).

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 16 November 2023. Downloaded on 11/18/2023 11:30:10 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

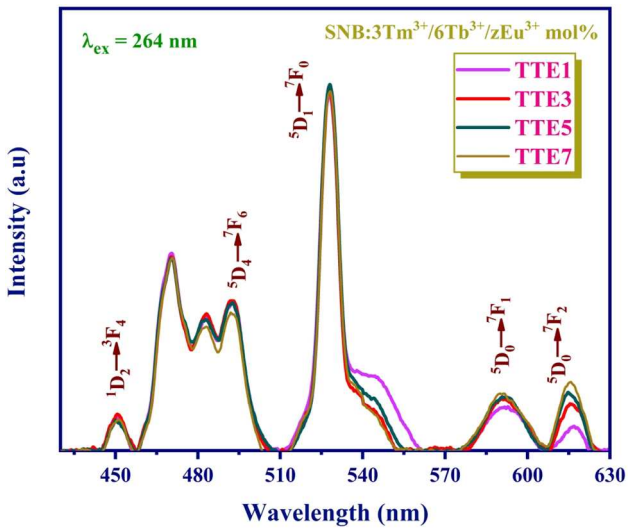


Fig. 9 PL spectra of SNB:3 mol% Tm^{3+}/6 mol% Tb^{3+}/*z* mol% Eu^{3+} phosphors under 264 nm excitation wavelength.

→ $^7F_5$ transition of Tb^{3+}. As the concentration of Tb^{3+} ions increases, the emission intensity of Tm^{3+} decline, while conversely, the emission intensity of Tb^{3+} and the energy transfer efficiency increases.

The emission spectra of SNB:3 mol% Tm^{3+}, 6 mol% Tb^{3+}, and *z* mol% Eu^{3+} phosphors illuminated by 264 nm UV light are depicted in Fig. 9. These spectra reveal distinct emission peaks associated with Tm^{3+}, Tb^{3+}, and Eu^{3+} ions at 450, 493, 528, 593, and 618 nm.[31,32] Notably, the positions of these characteristic peaks remain largely consistent across varying Eu^{3+} concentrations. Specifically, the emission intensity at 528 nm initially increases until the TTE5 phosphor, after which it diminishes with higher Eu^{3+} ion concentrations.

Fig. 10 provides further insight, illustrating changes in emission intensity. The emission intensity from Tm^{3+} ions at

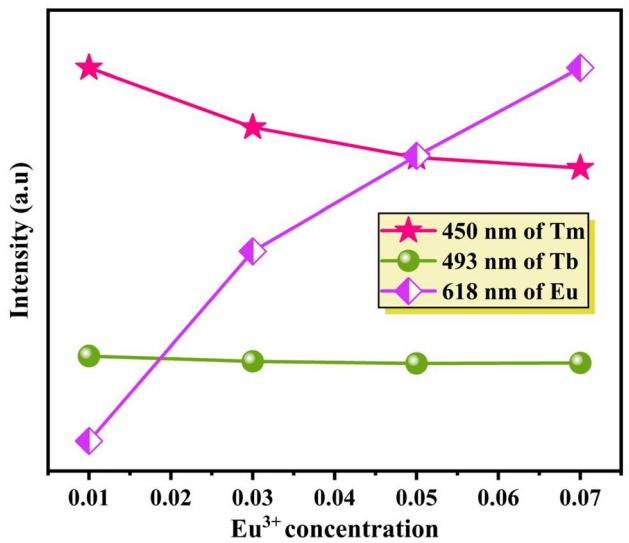


Fig. 10 Variation of the emission intensity of Tm^{3+}/Tb^{3+}/Eu^{3+} ions with the concentration of Eu^{3+} ions.

450 nm and Tb^{3+} ions at 493 nm declines, while the emission intensity attributed to Eu^{3+} ions at 618 nm rises as the Eu^{3+} concentration increases. This observation strongly implies a dual energy transfer mechanism, occurring from both Tm^{3+} and Tb^{3+} ions to Eu^{3+} ions.

Energy transfer mechanism of Tm^{3+}, Tb^{3+} and Eu^{3+} ions is illustrated in the schematic diagram presented in Fig. 11. The ions at ground state get excited to the higher energy levels of Tm^{3+}, Tb^{3+} and Eu^{3+} by the 264 nm UV excitation. After that ions gets de-excited to lower energy states $^5D_{1,0}$ of Eu^{3+} and $^5D_4$ of Tb^{3+} through non-radiative emission aided by phonons and relaxed to the ground state. In the case of Tm^{3+} ions, the blue photons emitted at 545 nm associated with $^1D_2 \rightarrow {}^3F_4$ transition. In the case of Tb^{3+} ions, the cyan and green photons emitted at 492 nm and 530 nm associated with $^5D_4 \rightarrow {}^7F_6$ transition and $^5D_4 \rightarrow {}^7F_5$ transition, respectively and for Eu^{3+} ions transitions at 593 nm ($^5D_0 \rightarrow {}^7F_1$) and 618 nm ($^5D_0 \rightarrow {}^7F_2$) emit orange and red photons, respectively. As the energy gap between the excited state $^1G_4$ of Tm^{3+} ions and $^5D_4$ of Tb^{3+} ions is smaller so they can bridge with the phonons of host lattice which results in non-radiative relaxation in Tm^{3+}. Some part of the energy is migrated to Tb^{3+} which results in transition from $^7F_6$ to $^5D_4$.[33] The non-radiative resonant energy transfer from Tm^{3+} ($^1G_4$) to Eu^{3+} ions ($^5D_2$) is because of lesser energy difference between these two states. This result in the Eu^{3+} ions transition from $^7F_1$ level to $^5D_2$ state and then return to $^5D_0$ state through non-radiatively which accompanied further radiative emission from this excited state to $^7F_1$ and $^7F_2$ state.

The energy transfer interactions between sensitizers and activators can be broadly categorized into two primary types: exchange interaction and electric multipolar interaction. The characterization of these interactions hinges on a critical parameter known as the critical distance (R_c) between the sensitizers and activators. If the value of R_c remains below 5 Å, exchange interaction is likely; however, if R_c surpasses this threshold, electric multipolar interaction becomes more probable.[34]

In order to determine the nature of the interaction between Tm^{3+} (sensitizer) and Tb^{3+} (activator), the critical distance R_c can be quantified using an equation outlined in existing literature.[35]

$$R_c = 2\left(\frac{3V}{4\pi N X_c}\right)^{\frac{1}{3}} \quad (3)$$

The computed R_c value is found to be 18.04 Å, which significantly exceeds the 5 Å threshold. This outcome strongly suggests that the energy transfer mechanism is primarily driven by electric multipolar interactions.

Utilizing Dexter's energy transfer formula for multipolar interaction and applying Reisfeld's approximation, the relationship can be presented as follows:[36,37]

$$\frac{I_{SO}}{I_S} \propto C_{Tm+Tb}^{\,n/3} \quad (4)$$

Here, I_{SO} and I_S denotes the luminescence intensities of Tm^{3+} ions without and with Tb^{3+} ions, respectively. C_{Tm+Tb} is the

Open Access Article. Published on 16 November 2023. Downloaded on 11/18/2023 11:30:10 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

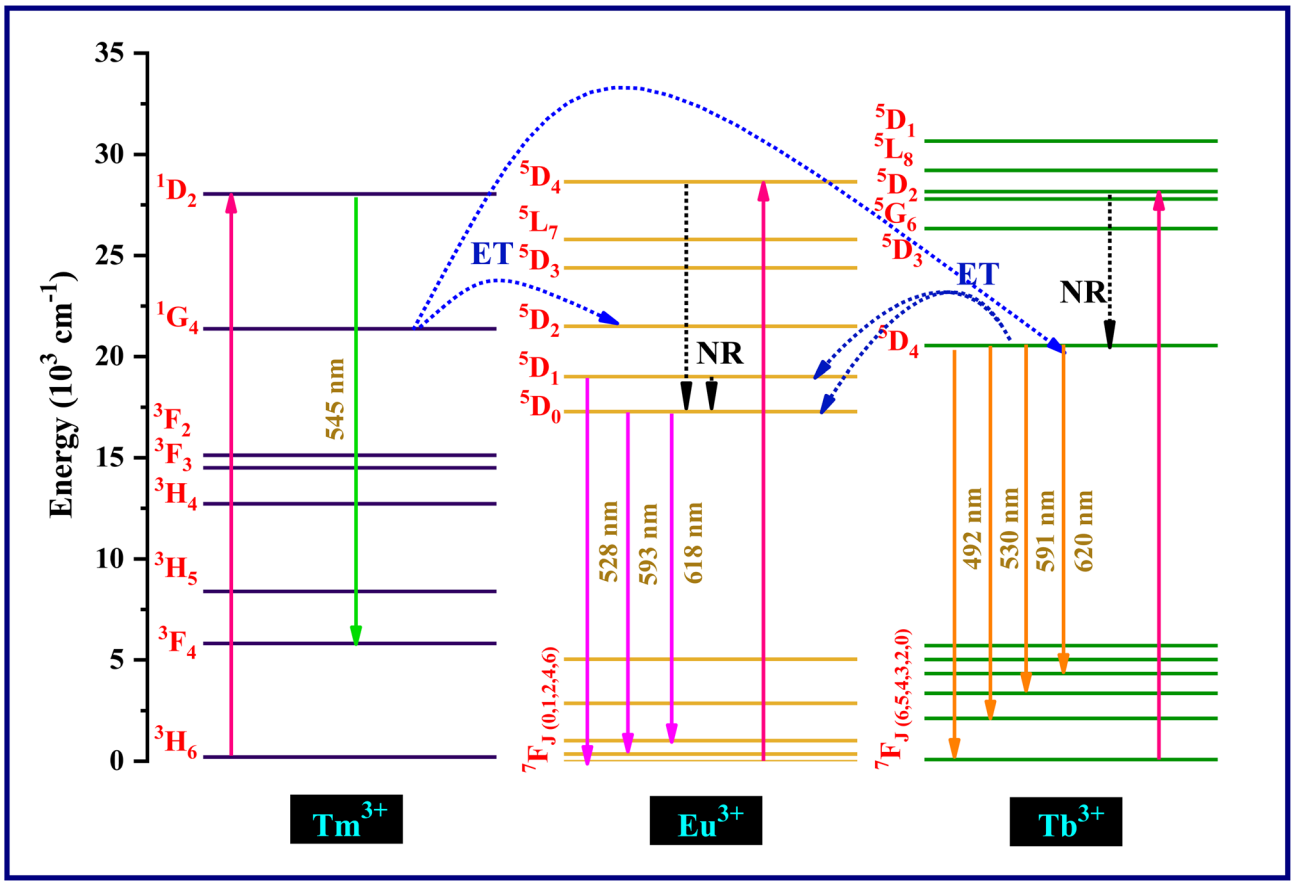


Fig. 11 Energy level diagram of Tm^{3+}/Tb^{3+}/Eu^{3+} ions at 264 nm.

combined concentration of Tb^{3+} and Tm^{3+} ions. The ratio $I_{SO}/I_S \propto C_{Tm+Tb}{}^{n/3}$ with $n = 6$, 8 and 10 corresponds to different exchange interactions: dipole–dipole, dipole–quadrupole, and quadrupole–quadrupole interactions, respectively.[38] The relationship between I_{SO}/I_S and $C_{Tm+Tb}{}^{n/3}$ is illustrated in Fig. 12. It is evident that the most optimal linear fitting is observed when

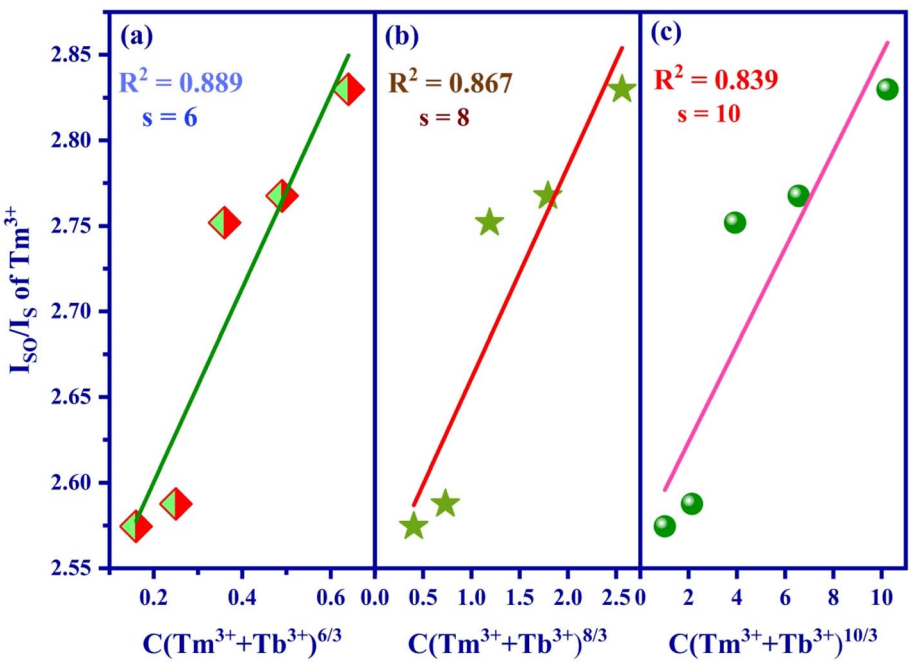


Fig. 12 Fitting relationship of I_{so}/I_s of Tm^{3+} on (a) $C\,(Tm^{3+} + Tb^{3+})^{6/3}$, (b) $C\,(Tm^{3+} + Tb^{3+})^{8/3}$ and (c) $C\,(Tm^{3+} + Tb^{3+})^{10/3}$.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 16 November 2023. Downloaded on 11/18/2023 11:30:10 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

$n = 6$. Consequently, the dominant mechanism for energy transfer from Tm^{3+} to Tb^{3+} is ascribed to dipole–dipole interactions.

To provide additional evidence for $Tm^{3+} \rightarrow Tb^{3+}$ energy transfer in SNB phosphor. Theoretically, it is known that when 3 mol% Tm^{3+} 6 mol% $Tb^{3+} \rightarrow z$ Eu^{3+} energy transfers, the time attenuation of Tb^{3+} ions will be accelerated energy transfer process 3 mol% Tm^{3+} 6 mol% $Tb^{3+} \rightarrow z$ Eu^{3+}. If 3 mol% Tm^{3+} 6 mol% Tb^{3+} and Eu^{3+} ions work independently without occurrence of 3 mol% Tm^{3+} 6 mol% $Tb^{3+} \rightarrow z$ Eu^{3+} energy transfers, the fluorescence lifetime of both remain equivalent to that observed in the single-doped sample. The experimental analysis of fluorescence lifetime of Eu^{3+} ions in SNB:3 mol% Tm^{3+}, 6 mol% Tb^{3+}, z Eu^{3+} phosphor was tested and the double exponential decay function was used to fit well:[39,40]

$$I(t) = I_0 + A_1 \exp\left(-\frac{t}{\tau_1}\right) + A_2 \exp\left(-\frac{t}{\tau_2}\right) \quad (5)$$

In eqn (5), I_t represents the luminous intensity at specific time t, I_0 is the background or detector zero offset. A_1 and A_2 are constants while τ_1 and τ_2 are the rapid and gradual life of exponential components, respectively. The computation of the

Table 1 Colorimetric parameters for SNB:$x$$Tm^{3+}$, SNB:3$Tm^{3+}$/$y$$Tb^{3+}$ and SNB:3Tm^{3+}/6Tb^{3+}/$z$$Eu^{3+}$ samples with excitation of 350 nm

	Color co-ordinates			
Sample	X	Y	CRI	CCT (K)
T1	0.288	0.172	—	6500
T3	0.271	0.157	—	3032
T5	0.279	0.163	—	3616
T7	0.280	0.165	—	4337
T10	0.293	0.176	—	7858
TT1	0.282	0.289	88.82	9677
TT2	0.293	0.303	89.41	8163
TT3	0.303	0.313	89.51	7256
TT4	0.302	0.314	89.52	7309
TT5	0.304	0.315	89.53	7165
TT6	0.298	0.309	89.48	7622
TT8	0.306	0.316	89.55	6984
TT10	0.302	0.312	89.45	7291
TTE1	0.299	0.307	89.34	7623
TTE3	0.298	0.303	88.87	7815
TTE5	0.301	0.307	89.11	7467
TTE7	0.299	0.303	88.64	7662

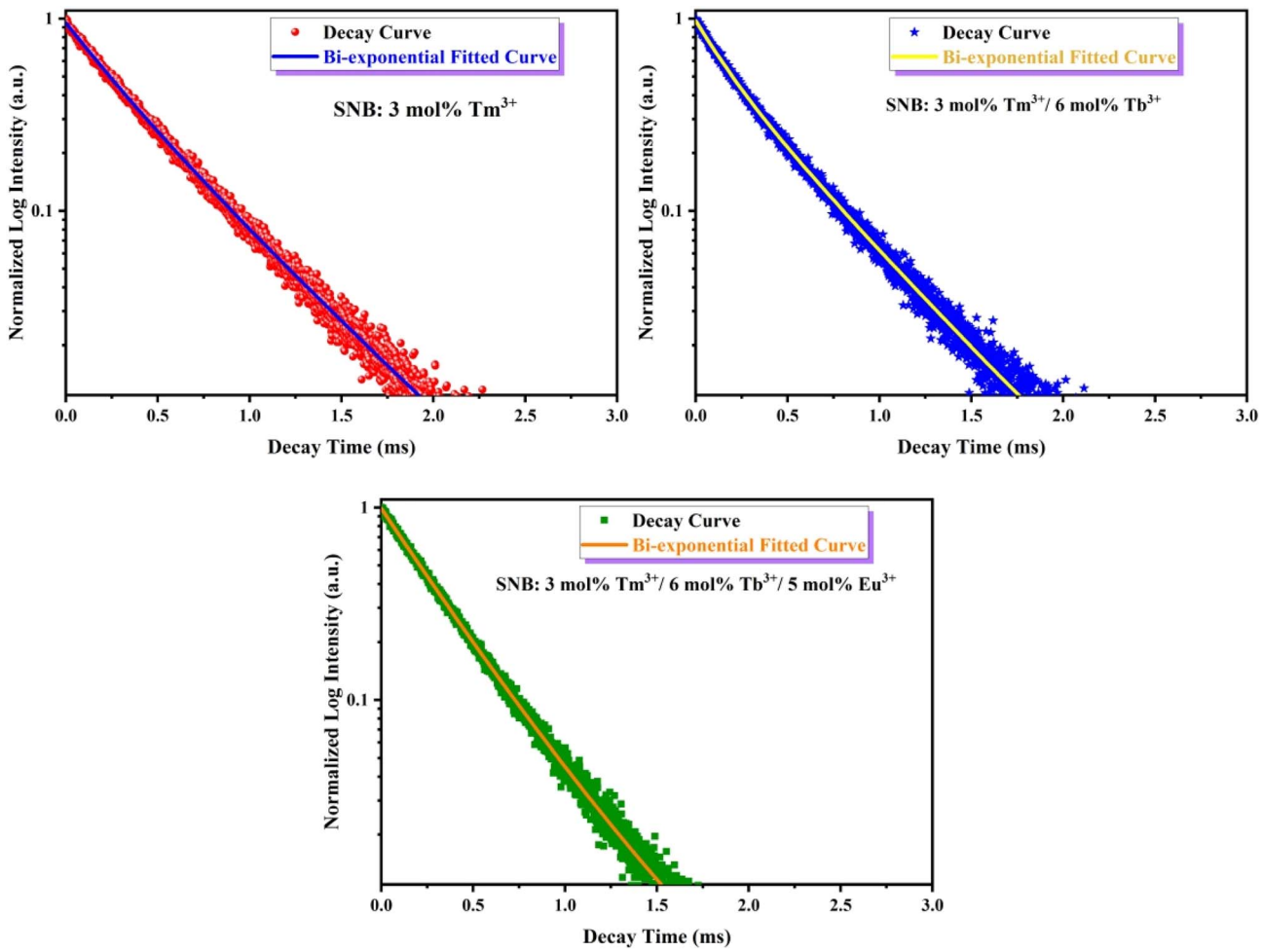


Fig. 13 PL decay curve for SNB:3 mol% Tm^{3+}, SNB:3 mol% Tm^{3+}/6 mol% Tb^{3+} and SNB:3 mol% Tm^{3+}/6 mol% Tb^{3+}/5 mol% Eu^{3+} phosphors.

Open Access Article. Published on 16 November 2023. Downloaded on 11/18/2023 11:30:10 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

average fluorescence lifetime (τ_{av}) can be performed using eqn (6):[41]

$$\tau_{avg} = \frac{A_1\tau_1^2 + A_2\tau_2^2}{A_1\tau_1 + A_2\tau_2} \quad (6)$$

The fluorescence lifetime decay curve of SNB:3 mol% Tm^{3+}, SNB:3 mol% Tm^{3+}, 6 mol% Tb^{3+} and SNB:3 mol% Tm^{3+}, 6 mol% Tb^{3+}, z Eu^{3+} phosphors is shown in Fig. 13 and the calculated τ_{av} value are 2.11, 1.90 and 1.64 ms, respectively. From the result it is evident that the decrement in the average life time of co doping sample SNB:3 mol% Tm^{3+}, 6 mol% Tb^3 is due to energy transfer between Tm^{3+} to Tb^{3+}. Further, the average lifetime decreased further when it is doped with Eu^{3+}, thus powerfully confirming the existence of 3 mol% Tm^{3+} 6 mol% Tb^{3+} → z Eu^{3+}.

The CIE chromaticity coordinates of SNB:$x$$Tm^{3+}$, SNB:3 mol% Tm^{3+}, $y$$Tb^{3+}$ and SNB:3 mol% Tm^{3+}, 6 mol% Tb^{3+}, $z$$Eu^{3+}$ phosphors are evaluated using formulation from the reported work in literature[42,43]

$$CCT = -449n^3 + 3525n^2 - 6823.3n + 5520.33 \quad (7)$$

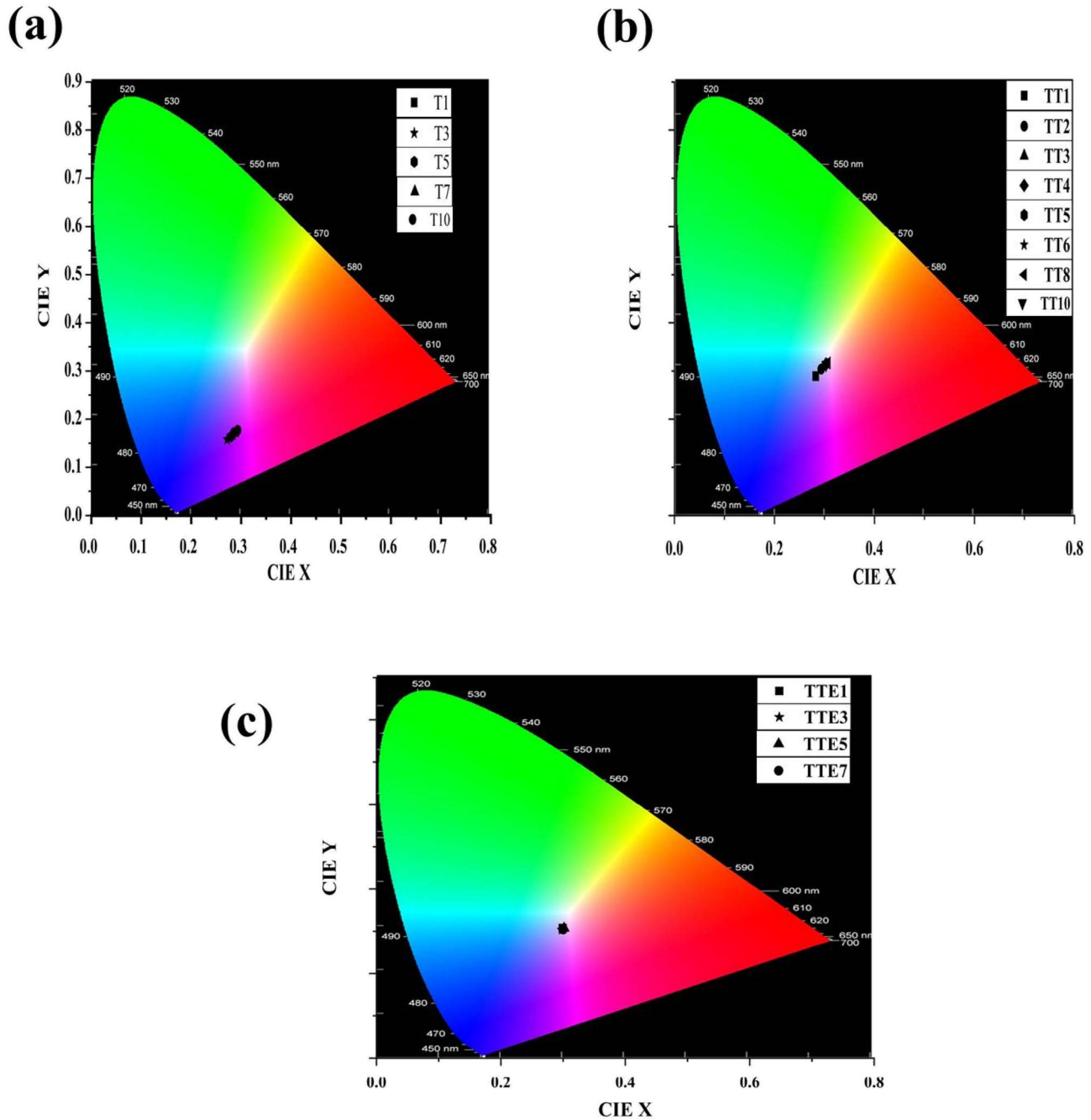


Fig. 14 CIE coordinates of (a) SNB:$x$$Tm^{3+}$, (b) SNB:3$Tm^{3+}$/$y$$Tb^{3+}$ and (c) SNB:3Tm^{3+}/6Tb^{3+}/$z$$Eu^3$ samples.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 16 November 2023. Downloaded on 11/18/2023 11:30:10 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

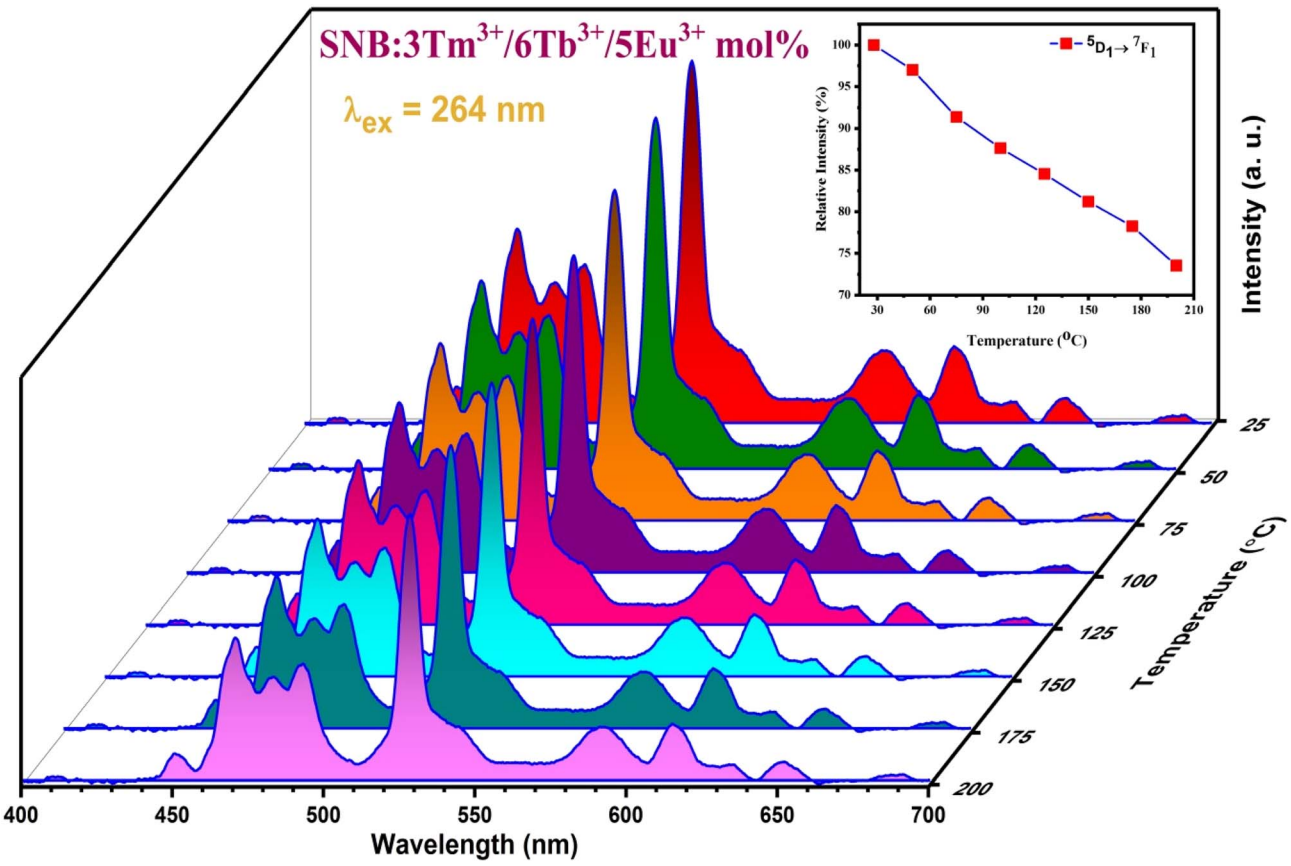


Fig. 15 TDPL spectra of SNB:3 mol% Tm^{3+}, 6 mol% Tb^{3+}, 5 mol% Eu^{3+} phosphor [inset exhibits the relative emission intensity variation for $^5D_1 \rightarrow {}^7F_1$ transition in the temperature range spanning from 28 to 200 °C].

and indicated in Table 1 and Fig. 14. These phosphors offer a remarkable avenue for color tunability, enabling the creation of customized illumination. The warm light can be achieved by adjusting the CCT and CRI values with NUV excitation of SNB:0.03Tm^{3+}, 0.06Tb^{3+}, zEu^{3+} phosphors. Notably, the CIE coordinates of TTE5 phosphor (0.3016, 0.3073) is close to standard value (0.31006, 0.3136). As highlighted in Table 1, increasing the concentration of Eu^{3+} concentration in the codoped phosphors has the effect of modifying the color temperature and enhancing the color rendering index by introducing the red emission into the spectra of white light. TTE1 phosphor represents the CCT of 7623 K and CRI 89.34 while the TTE3 phosphor emits the 7815 K CCT and 88.87 CRI white light. The result reveals that the TTE5 phosphor with higher Eu^{3+} concentration emits the cool light with CCT value of 7467 K and CRI value of 89.12. Hence, the white light with alterable CCT (7467–7815 K) and high CRI (>88) can be achieved for SNB:0.03Tm^{3+}, 0.06Tb^{3+}, zEu^{3+} phosphors with the control of Eu^{3+} concentration. Obviously, the TTE5 phosphor yield neutral white light with CCT of 7467 K and CRI of 89.12 with CIE coordinates (0.3016, 0.3073). These attributes position these phosphors as a strong contender for single-phased phosphors, holding significant promise for advancing neutral white light-emitting diodes (w-LEDs) with tunable color characteristics.

The thermal stability of as-prepared phosphors is crucial parameter in the application of W-LEDs device as operating temperature may increase significantly after the constant working of device at high temperature. The temperature dependent photoluminescence of SNB:3 mol% Tm^{3+}, 6 mol% Tb^{3+}, 5 mol% Eu^{3+} phosphor under 263 nm excitation was conducted in the range spanning from room temperature (28 °C) to 200 °C is shown in Fig. 15. The results indicate that the emission intensities of the as-prepared sample decreases with the rise in temperature because of thermal quenching but position remains unaltered. This decrement is attributed to the non-radiative transitions, therefore rate of intensity reduction changes over different emission peaks. This characteristic maintains the color stability of the as-prepared phosphors which indicates their reliable performance. Thermal stability was determined through the ratio of emission intensity at particular temperature to that at room temperature. The emission trend of the transition $^5D_1 \rightarrow {}^7F_1$ is depicted in the inset of Fig. 15. The strength of intensity at 150 °C and 200 °C relative to that at room temperature comes out to be 81.19% and 73.51%, respectively which is higher than reported work in the literature.[28] The results indicate the excellent thermal stability of the as-prepared phosphor.

3.7. Conclusions

We have synthesized a series of single-phase SNB phosphors co-doped with Tm^{3+}, Tb^{3+}, and Eu^{3+} ions. XRD analysis of the SNB phosphor confirms the formation of cubic structure within the

Open Access Article. Published on 16 November 2023. Downloaded on 11/18/2023 11:30:10 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Fm3m space group and its match with standard data from the JCPDS file. The quantum efficiency of TT6 phosphor found to be 91.3%. The TDPL spectra reveal that the PL intensity persists to 81.19% at 150 °C in compare to room temperature. Under an excitation wavelength of 360 nm, Tm^{3+} single-doped SNB samples exhibit blue luminescence, characterized by an initial increase followed by a subsequent decrease in luminescence intensity. The maximum intensity is reached at a 3 mol% doping concentration. Additionally, the photoluminescence emission spectra of SNB phosphors co-doped with Tm^{3+} and Tb^{3+} ions, excited at 263 nm, reveal a declining intensity of Tm^{3+} emissions alongside an increasing intensity of Tb^{3+} emissions. This shift in the behavior of Tm^{3+} and Tb^{3+} ion intensities signifies the occurrence of energy transfer from Tm^{3+} to Tb^{3+} ions. Upon doping Eu^{3+} ions into the optimized co-doped sample (3 mol% Tm^{3+} and 6 mol% Tb^{3+}), both Tb^{3+} and Tm^{3+} ion intensities decrease as the concentration of Eu^{3+} ions increases. This variation in intensity is attributed to the energy transfer process from the optimized co-doped sample to Eu^{3+} ions. By carefully adjusting the doping concentrations of Tm^{3+}, Tb^{3+}, and Eu^{3+} ions to approximately 3 mol%, 6 mol%, and 5 mol%, respectively, the emission color can be fine-tuned to achieve white light emission.

Author contribution

Ravina: conceptualization, methodology, writing – original draft. Kanishk Poria: conceptualization, methodology, writing – original draft. Mukesh K Sahu: data curation, writing – review & editing. A. Kumar: formal analysis. Anu: formal analysis. Sajjan Dahiya: resources, formal analysis. Nisha Deopa: supervision, methodology, software, validation, writing – review & editing formal analysis. A. S. Rao: writing – review & editing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

One of the authors, Ms Anu is very much thankful to CSIR/UGC for the award of a Junior Research Fellowship (JRF) (UGC Ref. 1352: CSIR-UGC NET DEC. 2018) to her.

References

1 J. Zheng, X. Wu, Q. Ren, W. Bai, Y. Ren, M. Wang and O. u. Hai, Investigation of luminescence properties and energy transfer in Sm^{3+} and Eu^{3+} co-doped $Sr_3Y(BO_3)_3$ red phosphors, *Opt. Laser Technol.*, 2020, **122**, 105857.

2 W. Zhang, H. Shen, X. Hu, Y. Wang, J. Li, Z. J. Zhu, Z. You and C. Tu, Solid-state synthesis, structure and spectroscopic analysis of Dy:$CaYAl_3O_7$ phosphors, *J. Alloys Compd.*, 2019, **781**, 255–260.

3 Q. Ren, Y. Zhao, X. Wu and O. Hai, Tb^{3+}/Eu^{3+}/Tm^{3+} co-doped single-phase phosphors emitting close to standard white light, *Opt. Laser Technol.*, 2021, **136**, 106774.

4 H. Patnam, S. K. Hussain, L. K. Bharat and J. S. Yu, Synthesis and luminescence properties of Eu^{3+} and Dy^{3+} ions single and co-doped $Ba_2LaV_3O_{11}$ phosphors for white-light applications, *Dyes Pigm.*, 2019, **162**, 583–589.

5 D. Rajesh, K. Brahmachary, Y. C. Ratnakaram, N. Kiran, A. P. Baker and G.-G. Wang, Energy transfer based emission analysis of Dy $^{3+}$/Eu $^{3+}$ co-doped ZANP glasses for white LED applications, *J. Alloys Compd.*, 2015, **646**, 1096–1103.

6 J. Li, H. Yan and F. Yan, A novel high color purity blue-emitting phosphor: $CaBi_2B_2O_7$: Tm^{3+}, *Mater. Sci. Eng., B*, 2016, **209**, 56–59.

7 T. S. Atabaev, H. H. Thi Vu, H. K. Kim and Y. H. Hwang, The optical properties of Eu^{3+} and Tm^{3+} codoped Y_2O_3 submicron particles, *J. Alloys Compd.*, 2012, **525**, 8–13.

8 G. Zhu, Z. P. Ci, Y. R. Shi, M. D. Que, Q. Wang and Y. H Wang, Synthesis, Crystal Structure and Luminescence Characteristics of a Novel Red Phosphor $Ca_{19}Mg_2(PO_4)_{14}$:Eu^{3+} for Light Emitting Diodes and Field Emission Displays, *J. Mater. Chem. C*, 2013, **1**, 5960–5969.

9 A. G. Bispo Jr, S. A. M. Lima and A. M. Pires, Energy transfer between terbium and europium ions in barium orthosilicate phosphors obtained from sol-gel route, *J. Lumin.*, 2018, **199**, 372–378.

10 P. C. Ricci, M. Salis, R. Corpino, C. M. Carbonaro, E. Fortin and A. Anedda, A Kinetics Model for Tb^{3+} Recombinations in Low Doped Tb:$Lu_{1.8}Y_{0.2}SiO_5$ Crystals, *J. Appl. Phys.*, 2010, **108**, 043512.

11 W. Zhao, X. Feng and B. Fan, Novel color tunable phosphors $NaYGeO_4$: Tm^{3+},Tb^{3+},Eu^{3+} for ultraviolet excited white LEDs with good thermal stability, *J. Mater. Sci.: Mater. Electron.*, 2020, **31**, 14434–14442.

12 V. V. Atuchin, B. I. Kidyarov and N. V. Pervukhina, Shortest chemical bond length as a criterion for searching new noncentrosymmetric niobate and tantalite crystals with high optical nonlinearity, *J. Cryst. Growth*, 2005, **275**, 1941–1946.

13 V. V. Atuchin, J. C. Grivel, A. S. Korotkov and Z. Zhang, Electronic parameters of $Sr_2Nb_2O_7$ and chemical bonding, *J. Solid State Chem.*, 2008, **181**, 1285–1291.

14 A. K. Vishwakarma, K. Jha, M. Jayasimhadri, A. S. Rao, K. Jang, B. Sivaiah and D. Haranath, Red light emitting $BaNb_2O_6$:Eu^{3+} phosphor for solid state lighting applications, *J. Alloys Compd.*, 2015, **622**, 97–101.

15 R. Lohan, A. Kumar, M. K. Sahu, A. Mor, V. Kumar, N. Deopa and A. S. Rao, Structural, thermal, and luminescence kinetics of $Sr_4Nb_2O_9$ phosphor doped with Dy^{3+} ions for cool w-LED applications, *J. Mater. Sci.: Mater. Electron.*, 2023, **34**, 694.

16 L. Alexander and H. P. Klug, Determination of crystallite size with the x-ray spectrometer, *J. Appl. Phys.*, 1950, **21**, 137–142.

17 M. Sharma, D. Kumar, J. P. Gewali and A. Thakur, Yttrium doped strontium nano hexaferrites: Influence of yttrium on the structural and magnetic properties of sol-gel synthesized nanocrystalline strontium hexaferrite, *Int. J. Mod. Phys.: Conf. Ser.*, 2022, **2267**, 012024.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 16 November 2023. Downloaded on 11/18/2023 11:30:10 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

18 K. Poria, R. Lohan, S. Bhatia, A. Kumar, R. Singh, N. Deopa, R. Punia, J. S. Shahi and A. S. Rao, Lumino-structural properties of Dy^{3+} activated $Na_3Ba_2LaNb_{10}O_{30}$ phosphors with enhanced internal quantum yield for w-LEDs, *RSC Adv.*, 2023, **13**(17), 11557–11568.

19 F. A. Kerim, H. F. Aly and A. Agramy, Infrared absorption spectra of some lanthanide acetylacetonate complexes, *Proc. Indian Acad. Sci.*, 1977, **85**, 559–566.

20 M. Afqir, A. Tachafine, D. Fasquelle, M. Elaatmani, J. C. Carru, A. Zegzouti, M. Daoud, S. Sayouri, T. D. Lamcharfi and M. Zouhairi, Structural, electric and dielectric properties of Eu-doped $SrBi_2Nb_2O_9$ ceramics obtained by co-precipitation route, Process, *Appl. Ceram.*, 2012, **12**, 72–77.

21 I. P. Sahu, D. P. Bisen, N. Brahme and R. K. Tamrakar, Generation of white light from dysprosium-doped strontium aluminate phosphor by a solid-state reaction method, *J. Mater. Sci.: Mater. Electron.*, 2016, **45**, 2222–2232.

22 W. T. Carnall, P. R. Fields and K. Rajnak, Electronic Energy Levels in the Trivalent Lanthanide Aquo Ions. I. Pr^{3+}, Nd^{3+}, Pm^{3+}, Sm^{3+}, Dy^{3+}, Ho^{3+}, Er^{3+}, and Tm^{3+}, *J. Chem. Phys.*, 1968, **49**, 4424.

23 W. T. Carnall, P. R. Fields and K. Rajnak, Electronic Energy Levels of the Trivalent Lanthanide Aquo Ions. III. Tb^{3+}, *J. Chem. Phys.*, 1968, **49**, 4447.

24 M. Wang, Y. Zhang, Z. Hu and J. Zhang, A single-phased $Ca_4(PO_4)_2O$: Tm^{3+}, Tb^{3+}, Eu^{2+} white light-emitting phosphor with tunable emission color, *J. Lumin.*, 2020, **224**, 117301.

25 W. Xiulan, D. Liang, Y. Zheng, M. Pei, Q. Ren and O. Hai, $Tm^{3+} \rightarrow Tb^{3+}$ energy transfer induced color-tunable in double-doped $LiLaSiO_4$ phosphors, *J. Lumin.*, 2021, **235**, 118027.

26 N. Deopa and A. S. Rao, Spectroscopic studies of single near ultraviolet pumped Tb^{3+} doped Lithium Lead Alumino Borate glasses for green lasers and tricolour w-LEDs, *J. Lumin.*, 2018, **194**, 56–63.

27 R. Pillania, A. Prasad, P. Rohilla, A. Shandilya and A. S. Rao, Photoluminescence and energy transfer studies on Tm^{3+}/Dy^{3+}/Eu^{3+}doped borosilicate glasses for color tunability and warm white light generation, *J. Non-Cryst. Solids*, 2023, **606**, 122192.

28 X. Wu, W. Bai, O. Hai, Q. Ren, J. Zheng and Y. Ren, Tunable color of Tb^{3+}/Eu^{3+}/Tm^{3+}-coactivated $K_3La(PO_4)_2$ via energy transfer: A single-phase white-emitting phosphor, *Opt. Laser Technol.*, 2019, **115**, 176–185.

29 X. Wu, L. Du, Y. Zheng, M. Pei, Q. Ren and O. Hai, $Tm^{3+} \rightarrow Tb^{3+}$ energy transfer induced color-tunable in double-doped $LiLaSiO_4$ phosphors, *J. Lumin.*, 2021, **235**, 118027.

30 X. Liu, C. Chen, S. Li, Y. Dai, H. Guo, X. Tang, Y. Xie and L. Yan, Host-sensitized and tunable luminescence of $GdNbO_4$: Ln^{3+} (Ln^{3+}= Eu^{3+}/Tb^{3+}/Tm^{3+}) nanocrystalline phosphors with abundant color, *Inorg. Chem.*, 2016, **55**, 10383–10396.

31 Y. Chen, X. Liu, G. Chen, T. Yang and C. Yuan, Luminescent characteristics of Tm^{3+}/Tb^{3+}/Eu^{3+} tri-doped borophosphate glasses for LED applications, *J. Mater. Sci.: Mater. Electron.*, 2017, **28**, 5592–5596.

32 Z. Wang, H. Tang, J. Xie, X. Zhang, X. Liu, X. Mi, Q. Liu, L. Lu and Z. Bai, Tunable color and energy transfer of $Y_2Mg_3Ge_3O_{12}$: Tb^{3+}, Tm^{3+}, Eu^{3+} phosphors with excellent thermal stability, *J. Solid State Chem.*, 2022, **314**, 123371.

33 H. Thakur, A. K. Gathania, I. Kumar and R. K. Singh, Tunable white light photoluminescence of a single phase Tm^{3+}/Tb^{3+}/Eu^{3+} codoped $GdPO_4$ phosphor, *Mater. Today Commun.*, 2023, **36**, 106458.

34 G. Blasse and B. C. Grabmaier, *Energy Transfer in Luminescent Materials*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1994, pp. 91–107.

35 G. Blasse, Energy transfer in oxidic phosphors, *Philips Res. Rep.*, 1969, **24**, 131–144.

36 K. Li, M. Shang, H. Lian and J. Lin, Recent development in phosphors with different emitting colors via energy transfer, *J. Mater. Chem. C*, 2016, **24**, 5507–5530.

37 J. Xiong, Q. Meng and W. Sun, Luminescent properties and energy transfer mechanism from Tb^{3+} to Eu^{3+} in $CaMoO_4$:Tb^{3+},Eu^{3+} phosphors, *J. Rare Earths*, 2016, **34**, 251–258.

38 D. L. Dexter, A theory of sensitized luminescence in solids, *J. Chem. Phys.*, 1953, **21**, 836–850.

39 C. H. Huang, T.-W. Kuo and T.-M. Chen, Novel Red-Emitting Phosphor $Ca_9Y(PO_4)_7$:Ce^{3+},Mn^{2+} with Energy Transfer for Fluorescent Lamp Application, *ACS Appl. Mater. Interfaces*, 2010, **2**, 1395–1399.

40 N. Deopa, M. K. Sahu, P. R. Rani, R. Punia and A. S. Rao, Realization of warm white light and energy transfer studies of Dy^{3+}/Eu^{3+} co-doped Li_2O-PbO-Al_2O_3-B_2O_3 glasses for lighting applications, *J. Lumin.*, 2020, **222**, 117166.

41 Y. Zhu, Q. Meng, W. Sun and S. Lü, $NaLa(MoO_4)_2$: Sm^{3+}, Tb^{3+} phosphor: optical temperature sensing material with a wide change range of luminescence color, *J. Lumin.*, 2020, **218**, 116854.

42 A. Kumar, A. Mor, M. K. Sahu, R. Pillania, S. Dahiya, N. Deopa, A. Malik, R. Punia and A. S. Rao, Spectral characteristics of Tb^{3+} doped ZnF_2–K_2O–Al_2O_3–B_2O_3 glasses for epoxy free tricolor w-LEDs and visible green laser applications, *J. Lumin.*, 2022, **244**, 118676.

43 C. S. McCamy, Correlated color temperature as an explicit function of chromaticity coordinates, *Color Res. Appl.*, 1992, **17**, 142–144.

J. Inst. Eng. India Ser. C
https://doi.org/10.1007/s40032-024-01026-2

ORIGINAL CONTRIBUTION

Energy-Efficient Fuzzy Scheduling System for Crankcase Covers Manufacturing

Sumit Chawla[1] · **Ranganath M. Singari**[2]

Received: 11 July 2023 / Accepted: 11 January 2024
© The Institution of Engineers (India) 2024

Abstract Scheduling automotive part manufacturing is more arduous and complex. Getting an optimum schedule for automotive part manufacturing is a very back-breaker task. In today's scenario, every industry and research organization need an energy-efficient system to cope up with the global environment. In the present work, we proposed energy-efficient fuzzy scheduling system for crankcase manufacturing. This study consists of the development of an energy-efficient fuzzy inference system, and its results are validated by a fuzzy set approach. Fuzzy logic can consider multiple criteria and can deal with uncertainty also. This scheduling system will help for identifying job priorities and select the best alternative route with multi-criteria scheduling through fuzzy logic. Three inputs are selected for both job priorities and route selection. Inputs are further divided into three ranges for developing 27 rules in the fuzzy reasoning tool. Fuzzy logic provides a decision by a combination of the rules for selecting job priorities and route selection. This study is very useful for all automotive industry as well as research organizations.

Keywords Scheduling · Energy · Fuzzy logic · Crankcase cover · Manufacturing

Introduction

The industrial sector has a significant impact on global energy consumption, and this consumption has been increasing rapidly over the past few decades. Sustainable manufacturing practices are essential to reduce the impact of this sector on the environment and promote a more sustainable future. There are several ways that industries can adopt sustainable manufacturing practices. One approach is to increase energy efficiency by using energy-efficient equipment, optimizing processes, and implementing energy management systems. This can help reduce energy consumption and lower costs. Many industries have already started to invest in renewable energy to reduce their carbon footprint and promote sustainability. In addition to energy efficiency and renewable energy, industries can also focus on waste reduction, water conservation, and sustainable materials sourcing. By adopting sustainable manufacturing practices, organizations can not only reduce their environmental impact, but also benefit from cost savings and increased efficiency [1].

Nowadays, every organization is facing the problem that how to convert their manufacturing system into an energy-efficient manufacturing system so that they can reduce energy consumption and can have the least adverse effect on the environment. Most of the researchers have used a static scheduling model in production scheduling problems for reducing energy consumption [2].

Today's research is based on decreasing the energy consumption of producing processes on the device level and the product stage by using the optimum schedule [3, 4]. Because of global warming and higher energy cost, efficient manufacturing is the need of every manufacturer as well as research organization. Efficient machining and a machine tool are needed for reducing energy consumption

✉ Sumit Chawla
chawlasumit4@gmail.com

[1] Applied Science Department (Mechanical), Bharati Vidyapeeth's College of Engineering, Delhi 110063, India

[2] Mechanical, Production, and Industrial Engineering Department, Delhi Technological University, Delhi 110042, India

Published online: 17 February 2024

within organizations [5–7]. Almost 37% of the world's total energy is consumed by an industrial sector such as construction, machining, mining, and agriculture. The energy-saving techniques, policies, and several ways are discussed by E. A. Abdelaziz et al. in their research work [8].

This study is done for developing an energy-efficient fuzzy-based scheduling system for crankcase cover manufacturing. Many researchers have used the fuzzy methodology in various scheduling problems as discussed in the literature part. For the first time, fuzzy approaches are applied for job prioritization and route selection of crankcase covers in the automobile industry.

This paper is organized into several parts. Part 2 briefs the literature review and summary of various articles related to energy-efficient scheduling. Part 3 explains the fuzzy methodology used in this paper. Part 4 defines the problem with the assumptions. Part 5 involves the crankcase cover prioritization using fuzzy rules. Part 6 involves the route selection calculation using both fuzzy approaches. Part 7 describes the results and discussions. Finally, part 8 concludes the results with the future scope.

Literature Review

Today's energy-efficient scheduling is an indispensable need of many manufacturing companies. Gahm, Denz, Dirr, and Tuma [9] developed a research framework for energy-efficient scheduling. A framework to analyze energy consumption characteristics in machining manufacturing systems is given by Li et al. [10]. Energy-efficient scheduling problems (EESP) deals with the reduction in energy consumption parameter, and all these studies are deterministic [11]. Makespan and energy consumption are the main parameters in these studies which must be balanced to achieve an energy-efficient environment [12]. Most of the researchers have used different metaheuristic techniques to solve various scheduling problems. The most preferred metaheuristic techniques are fuzzy logic and genetic algorithm (GA) [11, 13–15]. Fuzzy logic best deals with data uncertainty and incomplete information [16]. Table 1 summarizes the various research and articles related to EESP. Energy efficiency barriers are generally four types such as economic non-market failure, economic market failure, behavioral barrier, and organizational barrier. [33].

Fuzzy schedule modeling has been successfully applied in manufacturing [32, 34], project scheduling [35], vehicle routing problems [36], timetabling scheduling problems [37], and several other scheduling applications, e.g., casting scheduling problems of steelmaking [19] and nurse scheduling problems in hospitals [38]. Zhang et al. [35] analyzed the real-world project scheduling problems using the fuzzy programming models considering the total crash cost as the objective function, and parameters include fuzzy duration, fuzzy crash time, and fuzzy completion time with possibility level α.

Past studies show that most researchers have successfully applied fuzzy approaches for solving various scheduling problems. After reviewing the existing literature, it is found that the route selection and job prioritization of crankcase cover manufacturing in the automobile industry are an untouched area of research. The above studies also show that fuzzy-based methodologies are effective for the generation of optimum schedules in various industries. Therefore, the present study initially aims for job prioritization and route selection of the crankcase cover manufacturing using the Mamdani fuzzy inference system. Later, the fuzzy set approach is applied for the validation of results.

Fuzzy Methodology

Fuzzy logic is a mathematical framework for dealing with uncertainty and imprecision, and it was developed by Lotfi Zadeh in 1965. He showed this approach as a way of processing data by providing a piece of incomplete and uncertain information rather than crisp information [39]. Most of the control system uses a fuzzy logic control-based methodology. Fuzzy rules used in these systems hold the performance data [40].

Steps in the fuzzy methodology are shown as a flowchart in Fig. 1a. Step 1 includes the selection of input and output variables. Step 2 consists of the selection of the membership function for these variables. Step 3 includes the development of a linguistic rule base based on the performance data gathered from the various experts. Step 4 includes the defuzzification process in which linguistic variables are converted into a crisp form.

Route selection can be done in two ways in the fuzzy logic system. These two ways are the fuzzy set approach and the fuzzy reasoning tool in MATLAB.

Fuzzy Set Approach

The route can be optimized by minimizing several factors. These factors are travel time, work in the queue, total processing time, energy consumption, number of processing steps, and number of machines blocked [41].

Among these factors, we have used only four factors (travel time, work in queue, total processing time, and energy consumption) in our study. This factor's contributions in terms of membership function are defined in Eqs. (1–4). A membership function $u_1(i)$ for the minimization of the total processing time (goal 1) is defined by Eq. (1).

Table 1 Summary of conducted research on EESP

Authors	Approaches	Techniques	Scheduling Type	Parameter optimized
Wang et al. [11]	Metaheuristic Approach	Fuzzy Logic	Batch scheduling	Energy consumption, tardiness
Aghelinejad et al. [13]	Metaheuristic Approach	GA	Production scheduling	Energy cost
Yan et al., [14]	Metaheuristic Approach	GA	Flexible flow shop scheduling	Energy consumption, Makespan, Cutting energy, Cutting times
May et al. [15]	Metaheuristic Approach	GA	Job shop scheduling (JSS)	Energy consumption, Makespan
Tang et al. [2]	Metaheuristic Approach	PSO	Flexible flow shop Scheduling	Energy consumption, Makespan
Fang & Lin [17]	Metaheuristic Approach	PSO	Parallel-machine scheduling	Tardiness penalty and power cost
Garcia-Santiago et al. [18]	Metaheuristic Approach	Harmony Search	Production scheduling	Energy consumption
Dorry [19]	Hybrid/Cluster Approach	Hybrid Fuzzy PSO Approach	Steelmaking- Continuous Casting Scheduling	Waiting time of charges
Lu, Gao, Li, Pan, & Wang [20]	Hybrid/Cluster Approach	Hybrid backtracking search algorithm	Flow shop scheduling	Energy consumption, Makespan
Liu, Lohse, Petrovic, & Gindy [21]	Hybrid/Cluster Approach	Non-dominant Sorting GA	JSS	Total energy consumption and total weighted tardiness
Mokhtari & Hasani [22]	Hybrid/Cluster Approach	Evolutionary Algorithm	JSS	Total energy cost, Total completion time, Total availability
Jiang, Zuo, & Mingcheng [23]	Operation Research	Non-dominant Sorting GA	JSS	Energy consumption, Makespan, Processing cost
Wu & Sun [24]		Green Scheduling Algorithm		Energy Consumption, Makespan, No. of turning on–off machines
Gong, De Pessemier, Joseph, & Martens [25]	Operation Research	Mix-integer programming	Production Scheduling	Energy consumption, Demand response
Bruzzone, Anghinolfi, Paolucci, & Tonelli [26]	Operation Research	Operation Scheduling approach	Flexible flow shops	Total tardiness, Makespan
L. Li, Huang, Zhao, & Liu [27]	Operation Research	Operation Scheduling approach	Operation scheduling	Energy consumption, Makespan
Chen, Zhang, Arinez, & Biller [28]	Statistical Approach	Markovian analysis	Operation scheduling	Energy consumption, Productivity
Choi, Y.C [29]	Algorithm-Based Approach	Dispatching rule-based scheduling algorithms	Single Machine Scheduling	Average Energy consumption, Mean tardiness
Lee S et al. [30]	Algorithm-Based Approach	Dynamic control algorithm	Single Machine Scheduling	Energy consumption, Total penalty cost
Merkert et al. [31]	Literature review papers	Review Paper	EESP	Energy consumption, Energy Efficiency, Profit
Duflou et al. [32]	Literature review papers	Review Paper	EESP	Energy Efficiency, Resource efficiency
Gahm et al. [9]	Literature review papers	Review Paper	EESP	Energy demand, Energy supply, Energetic Coverage

$$u_1(i) = \frac{(\mathrm{TPT}_b - \mathrm{TPT}_i)}{(\mathrm{TPT}_b - \mathrm{TPT}_a)} \tag{1}$$

where TPT_i = total processing time required for routing i, and $\mathrm{TPT}_b - \mathrm{TPT}_a$ represents the difference between the maximum and minimum total processing time. Similarly, the membership functions for the minimization of total work in queue (goal 2), total travel time (goal 3), and total energy consumption (goal 4) are defined by Eqs. (2, 3 and 4), respectively. The formula for the calculation of the final membership function $u_a(i)$ is given in Eq. (5).

$$u_2(i) = \frac{(\mathrm{TWIQ}_b - \mathrm{TWIQ}_i)}{\mathrm{TWIQ}_b} \tag{2}$$

$$u_3(i) = \frac{(\mathrm{TTT}_b - \mathrm{TTT}_i)}{(\mathrm{TTT}_b - \mathrm{TTT}_a)} \tag{3}$$

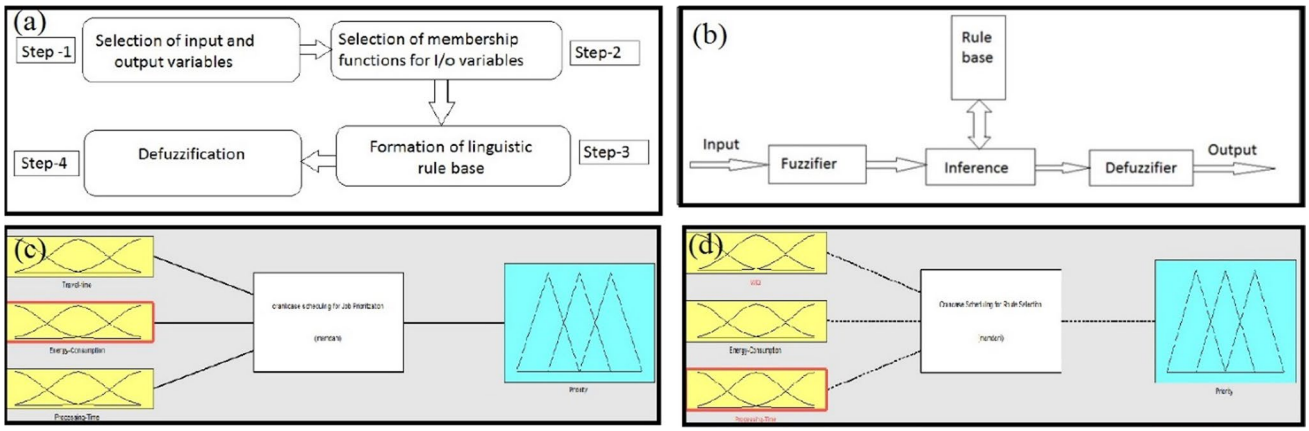


Fig. 1 **a** Flowchart of steps in the fuzzy methodology, **b** Structure of fuzzy reasoning tool [43], **c** Structure of Mamdani fuzzy rule-based system for evaluating job prioritization **d** Structure of Mamdani fuzzy rule-based system for route selection

$$u_4(i) = \frac{(\mathrm{TEC}_b - \mathrm{TEC}_i)}{(\mathrm{TEC}_b - \mathrm{TEC}_a)} \tag{4}$$

$$u_a(i) = \sum_{j=1}^{n} \left[\frac{w_j}{\sum_{j=1}^{n} w_j} \times u_j(i) \right] \tag{5}$$

where $u_j(i) =$ membership of routing i to goal j and $w_j =$ weight of goal j in Eq. (5).

Fuzzy Reasoning Tool in MATLAB

Fuzzy logic gives the best results in cases of uncertainty and any missing information. This technique can be effectively applied to all control problems [42]. The structure of a fuzzy reasoning tool typically includes a fuzzifier, a rule base, an inference engine, and a defuzzifier. The fuzzifier converts the crisp input into fuzzy input, the inference engine applies the rules in the rule base to the fuzzy input to generate fuzzy output, and the defuzzifier converts the fuzzy output into crisp output. This structure allows the system to handle imprecise and uncertain information, making it useful in a wide range of applications where traditional logic systems may not be effective. The structure of the fuzzy reasoning tool is shown in Fig. 1b [43].

Development of Mamdani Fuzzy Scheduling System

The Mamdani fuzzy scheduling system is a type of fuzzy reasoning tool that is used to optimize scheduling decisions in a manufacturing environment. It is named after its creator, Lotfi A. Zadeh's student, Ebrahim Mamdani. Mamdani fuzzy scheduling system is used to estimate job prioritization and route selection. The set of output data is obtained through the given input condition in the system. The proposed Mamdani fuzzy scheduling system for job prioritization and route selection is presented in Fig. 1c, d.

The input variables for job prioritization are identified as processing time, travel time, and energy consumption, and the output variable is the priority. The input variables for route selection are identified as work in queue, processing time, and energy consumption, and the output variable is the priority. The triangular membership function is used for defining the output and input variables.

In the fuzzy reasoning tool, the inputs cycle time, due date, and energy consumption have three membership functions each, and each input is further divided into three states low, medium, and high. Thus, a total of 27 rules can be made for job prioritization. Similarly, the inputs work in queue, cycle time, and energy consumption for route selection have three membership functions, and they are also divided into three states low, medium, and high. Thus, a total of 27 rules can also be made for route selection.

Defuzzification is the process of linguistic values into crisp values. There are commonly three techniques used for defuzzification as described as follows [44]:

(a) Center of gravity method (COG)
(b) Mean of maximum method (MOM)
(c) Height method

Problem Definition

Our main problem is how to get an optimum schedule for crankcase cover manufacturing. For obtaining the optimum schedule, we have used the fuzzy inference system and the fuzzy set approach. This study consists of the scheduling of four crankcase covers as shown in Fig. 2a-d.



Fig. 2 Four crankcase covers **a** Cover left side crankcase KWPG, **b** Cover left crankcase K38, **c** Cover crankcase 206 G **d** Cover right crankcase KTE

Assumptions Made in the Present Work

The following assumptions were made:

1. Setup times are assumed to be constant.
2. It is assumed that all the tools are available when needed.
3. Dynamic uncertainty is not taken into consideration, i.e., the machines are not subject to any failure.
4. Each workstation can process only one job at a time.
5. There can be multiple routes possible for the jobs for sending them to different workstations, and each route may have a different cycle time.
6. Delays in accessing any set of information are assumed to be negligible.
7. The speed of all the laborers is supposed to be constant, i.e., it is independent of the weight and size of jobs.
8. Orders arrive in the system randomly. So, crankcase cover manufacturing processes are assumed Poisson-distributed.

The input variables used in the fuzzy reasoning tool in MATLAB to identify the job priority are:

(a) Processing time (20–40 min)
(b) Travel time (2 to 10 min)
(c) Energy consumption (3 to 6 KW)

Crankcase Cover Prioritization Using the Fuzzy Reasoning Tool

All the inputs and their corresponding ranges are entered in the fuzzy reasoning tool of MATLAB. These inputs are entered in the FIS editor. In the FIS editor, we can choose the defuzzification method. For this scheduling problem, the centroid defuzzification method is selected in the FIS editor. All input variables for job priority are divided into three parts: small, medium, and high.

Fuzzy Rules for Job Prioritization

Travel time, energy consumption, and processing time are the three main inputs/factors responsible for job prioritization. The output variable 'priority' is divided into nine minimum categories (MIN), negative low (NLO), low (LO), negative average (NAV), average (AV), positive average (PAV), high (HI), positive high (PHI), and maximum (MAX) [45]. The fuzzy rules for job prioritization using these three inputs and one output (priority values) are given in Table 2. These rules are generated with the help of the literature review and expert views.

Table 2 Fuzzy rule table for job prioritization

Travel time	Energy consumption			Processing time
	High	Medium	Small	
Small	HI	PHI	MAX	Small
Medium	PAV	HI	PHI	Medium
High	MIN	NLO	NAV	High
Small	PAV	HI	PHI	Small
Medium	AV	PAV	HI	Medium
High	NLO	LO	NAV	High
Small	AV	PAV	HI	Small
Medium	NAV	AV	PAV	Medium
High	LO	NAV	AV	High

Route Selection Calculation Using Fuzzy Logic

Route selection can be done in two ways. These two ways are as follows:

1. Calculation and analysis using fuzzy reasoning tool
2. Calculation using the fuzzy set approach

Using Fuzzy Reasoning Tool

The fuzzy set approach is a simple calculation-based approach using the contribution of goals in terms of membership functions. We can do routing by using fuzzy reasoning tool in MATLAB which takes less time than the fuzzy set approach. MATLAB approach is difficult to understand in comparison with the fuzzy set approach. MATLAB approach is simple if an equal contribution of inputs is taken. But it becomes a complex approach if we vary the contribution of inputs.

Fuzzy Rules for Route Selection

Among the three variables for job prioritization, two variables are the same for route selection (energy consumption and processing time), and the third variable is taken works in the queue. Their fuzzy rule table is given in Table 3.

Table 3 Fuzzy rules for route selection of jobs

Work in queue	Energy consumption			Processing time
	High	Medium	Small	
Small	NAV	NAV	NHI	Small
Medium	NLO	NAV	AV	Medium
High	MIN	NLO	NAV	High
Small	AV	HI	PHI	Small
Medium	LO	NAV	NAV	Medium
High	NLO	NAV	PAV	High
Small	HI	PHI	MAX	Small
Medium	PAV	HI	PHI	Medium
High	AV	PAV	HI	High

Route Time Calculation for All Jobs

We have taken four crankcase cover on six machine scheduling problems. Some data have been collected from the industry and some from the literature review. The processing times of various crankcase covers on different machines are shown in Table 4.

Fettling is the most ignored process in the casting process. Fettling can be done manually (files, chisel, or hammer) or with some automatic CNC machine. Large castings like cylinder head or cylinder block require a CNC machine for fettling, whereas small and medium casting like crankcase cover for two-wheeler vehicles can be fettled manually. The route time of all four crankcase cover on six machines is given in Table 5.

Route Selection Using Fuzzy Set Approach

Each goal plays a significant role in finding the optimized route. Its contribution in terms of the membership function is shown in Table 6. The maximum value of $u_j(i)$ corresponds to the maximum value of $u_a(i)$ shown in Table 7.

Results & Discussion

A rule viewer in fuzzy reasoning tool typically displays the rules in a graphical format that allows users to easily visualize the relationships between inputs and outputs. The viewer may display the rules as a set of if–then statements or in a more visual format, such as a graph or a table. Some rule viewers may also allow users to edit or modify the rules directly, which can be useful for tuning and optimizing the performance of a fuzzy reasoning tool. The rule viewer for crankcase cover prioritization is shown in Fig. 3.

Impacts of All Inputs on Job Prioritization and Route Selection

A surface viewer in fuzzy reasoning tool is a graphical interface examines the output surface of a fuzzy scheduling system for any one or two inputs. The surface viewer for crankcase cover prioritization is shown in Fig. 4a-c

Table 4 Processing time (mins) of various crankcase covers on different machines

M/c	KWPG	K38	206G	KTE
Cold chambers die casting machine (DCM)	1.5	1.3	1.5	1.7
Manual fettling (MF)/CNC fettling (CF)	0.18/0.54	0.17/0.51	0.18/0.54	0.20/0.60
Drilling machine (DM)	0.25	0.15	0.40	0.50
Vertical milling center (VMC)	3.5	2	4	6
Polishing machine/Buffing machine (BM)	0.22	0.21	0.22	0.23
Surface treatment/Painting shop (PS)	20	19.15	22.7	23.4

Table 5 Route times for various crankcase cover

Route times of crankcase cover KWPG				
Route (M/C sequence)	Work in queue	Processing time (mins)	Travel time (mins)	Energy consumption (KW)
DCM-MF-DM-VMC-BM-PS	10	25.65	6.4	5.1
DCM-CF-DM-VMC-BM-PS	8	26.01	6.5	5.6
DCM-VMC-DM-VMC-BM-PS	5	28.97	7.2	5.3
Route times of crankcase cover K38				
DCM-MF-DM-VMC-BM-PS	9	23	5.7	4.6
DCM-CF-DM-VMC-BM-PS	7	23.34	5.8	5.2
DCM-VMC-DM-VMC-BM-PS	4	24.83	6.2	4.9
Route times of crankcase cover 206G				
DCM-MF-DM-VMC-BM-PS	13	29	7.2	5.3
DCM-CF-DM-VMC-BM-PS	11	29.36	7.3	5.7
DCM-VMC-DM-VMC-BM-PS	8	32.82	8.2	5.5
Route times of crankcase cover KTE				
DCM-MF-DM-VMC-BM-PS	15	32.03	8	5.6
DCM-CF-DM-VMC-BM-PS	13	32.43	8.1	5.9
DCM-VMC-DM-VMC-BM-PS	11	37.83	9.4	5.8

Table 6 Membership function values

Results corresponding to goal 1				Results corresponding to goal 2			
$[u_1(1)]_{KWPG}$	0.717	$[u_1(1)]_{K38}$	0.85	$[u_2(1)]_{KWPG}$	0.625	$[u_2(1)]_{K38}$	0.562
$[u_1(2)]_{KWPG}$	0.699	$[u_1(2)]_{K38}$	0.833	$[u_2(2)]_{KWPG}$	0.5	$[u_2(2)]_{K38}$	0.437
$[u_1(3)]_{KWPG}$	0.551	$[u_1(3)]_{K38}$	0.758	$[u_2(3)]_{KWPG}$	0.312	$[u_2(3)]_{K38}$	0.25
$[u_1(1)]_{206G}$	0.55	$[u_1(1)]_{KTE}$	0.398	$[u_2(1)]_{206G}$	0.812	$[u_2(1)]_{KTE}$	0.937
$[u_1(2)]_{206G}$	0.532	$[u_1(2)]_{KTE}$	0.378	$[u_2(2)]_{206G}$	0.687	$[u_2(2)]_{KTE}$	0.812
$[u_1(3)]_{206G}$	0.359	$[u_1(3)]_{KTE}$	0.108	$[u_2(3)]_{206G}$	0.5	$[u_2(3)]_{KTE}$	0.687
Results corresponding to goal 3				Results corresponding to goal 4			
$[u_3(1)]_{KWPG}$	0.55	$[u_3(1)]_{K38}$	0.462	$[u_4(1)]_{KWPG}$	0.7	$[u_4(1)]_{K38}$	0.533
$[u_3(2)]_{KWPG}$	0.562	$[u_3(2)]_{K38}$	0.475	$[u_4(2)]_{KWPG}$	0.867	$[u_4(2)]_{K38}$	0.733
$[u_3(3)]_{KWPG}$	0.65	$[u_3(3)]_{K38}$	0.525	$[u_4(3)]_{KWPG}$	0.767	$[u_4(3)]_{K38}$	0.633
$[u_3(1)]_{206G}$	0.65	$[u_3(1)]_{KTE}$	0.75	$[u_4(1)]_{206G}$	0.767	$[u_4(1)]_{KTE}$	0.867
$[u_3(2)]_{206G}$	0.662	$[u_3(2)]_{KTE}$	0.762	$[u_4(2)]_{206G}$	0.9	$[u_4(2)]_{KTE}$	0.967
$[u_3(3)]_{206G}$	0.775	$[u_3(3)]_{KTE}$	0.925	$[u_4(3)]_{206G}$	0.833	$[u_4(3)]_{KTE}$	0.933

Table 7 Maximum value of $u_a(i)$ for different crankcase cover

Route (M/C sequence)	KWPG	K38	206G	KTE
DCM-MF-DM-VMC-BM-PS	0.648	0.602	0.695	0.738
DCM-CF-DM-VMC-BM-PS	0.657	0.619	0.695	0.729
DCM-VMC-DM-VMC-BM-PS	0.57	0.541	0.617	0.663

and for route selection of crankcase cover is shown in Fig. 4d-f.

Figure 4a shows higher priority values of job prioritization for low values of energy consumption and travel time. This diagram also shows that energy consumption has an inverse relation with priority, whereas the priority is high for low values of travel time, and it remains constant for a particular range and then suddenly decreases for the high value of travel time. In Fig. 4b, we can observe the moderate values of priority for high values of travel time and processing time and high values of priority for low values of travel time and processing time. Travel time and processing time

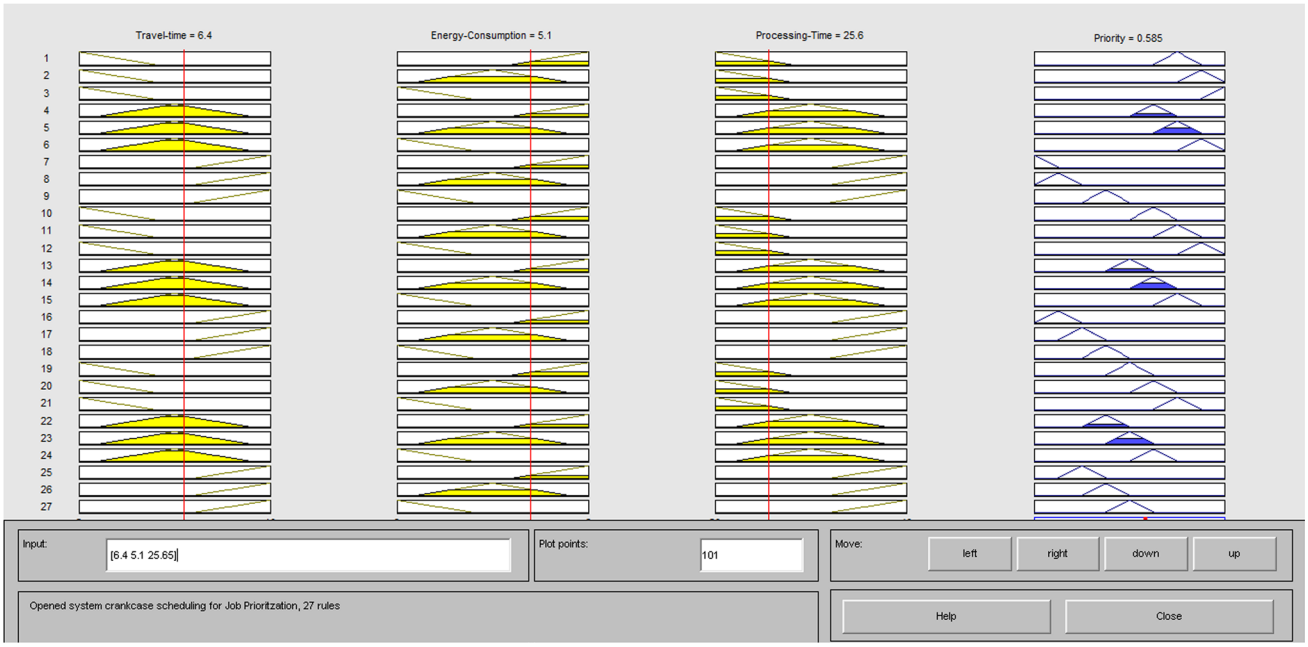


Fig. 3 Rule viewer window for crankcase cover prioritization

have an almost similar impact on priority. Figure 4c depicts the impacts of processing time and energy consumption on priority, and this variation is like Figure 4a. Figure 4d, e describes the scheduling parameter's impact on the route selection. Travel time and WIQ effects on the priority show a similar variation with processing time and WIQ on the priority. Figure 4f depicts the high values of route selection priority on the low values of processing time and travel time and moderate values of priority on the high values of processing time and travel time.

Route Priority for All Jobs Using Fuzzy Reasoning Tool

The route priority of four crankcase covers obtained by fuzzy reasoning tool is given in Table 8.

Results Obtained Using Fuzzy Set Approach

This approach is based on the calculation of membership function values. Based on the maximum final membership function values, route selection is done which is shown in Table 9. The final sequence obtained by the fuzzy reasoning tool and fuzzy set approach is shown in Table 10.

For KWPG, K38, and 206G, the maximum final membership values for both routes (DCM-MF-DM-VMC-BM-PS, DCM-CF-DM-VMC-BM-PS) are approximately the same. So, both sequences are shown as optimum routes in Table 9. Some of the sequences are shown in bold because these are the same sequences that we get using a fuzzy reasoning tool in MATLAB which is verifying our result. All maximum priority value is obtained corresponding to the machine sequence (DCM-MF-DM-VMC-BM-PS).

Conclusion

It is observed the following remarkable conclusions from this study with some future research expansion.

- The study described focuses on using a fuzzy reasoning tool to address the energy-efficient crankcase cover process scheduling problem under uncertain processing times.
- The fuzzy reasoning tool was used to prioritize jobs and select routes for manufacturing, and a Mamdani fuzzy scheduling system was developed to provide a trade-off between energy consumption and makespan.
- Results show the industry must use manual fatling operation rather than CNC or VMC machines for energy-efficient scheduling. Job priority sequence obtained by both fuzzy techniques for four crankcase cover problem = {K38, KWPG, 206G, KTE}.
- The results were validated using a fuzzy set approach, making them valuable for production engineers and supervisors in automotive industries.
- In terms of future research, the study suggests the application of another metaheuristic approach such as GA or PSO to the same problem with variable setup time and

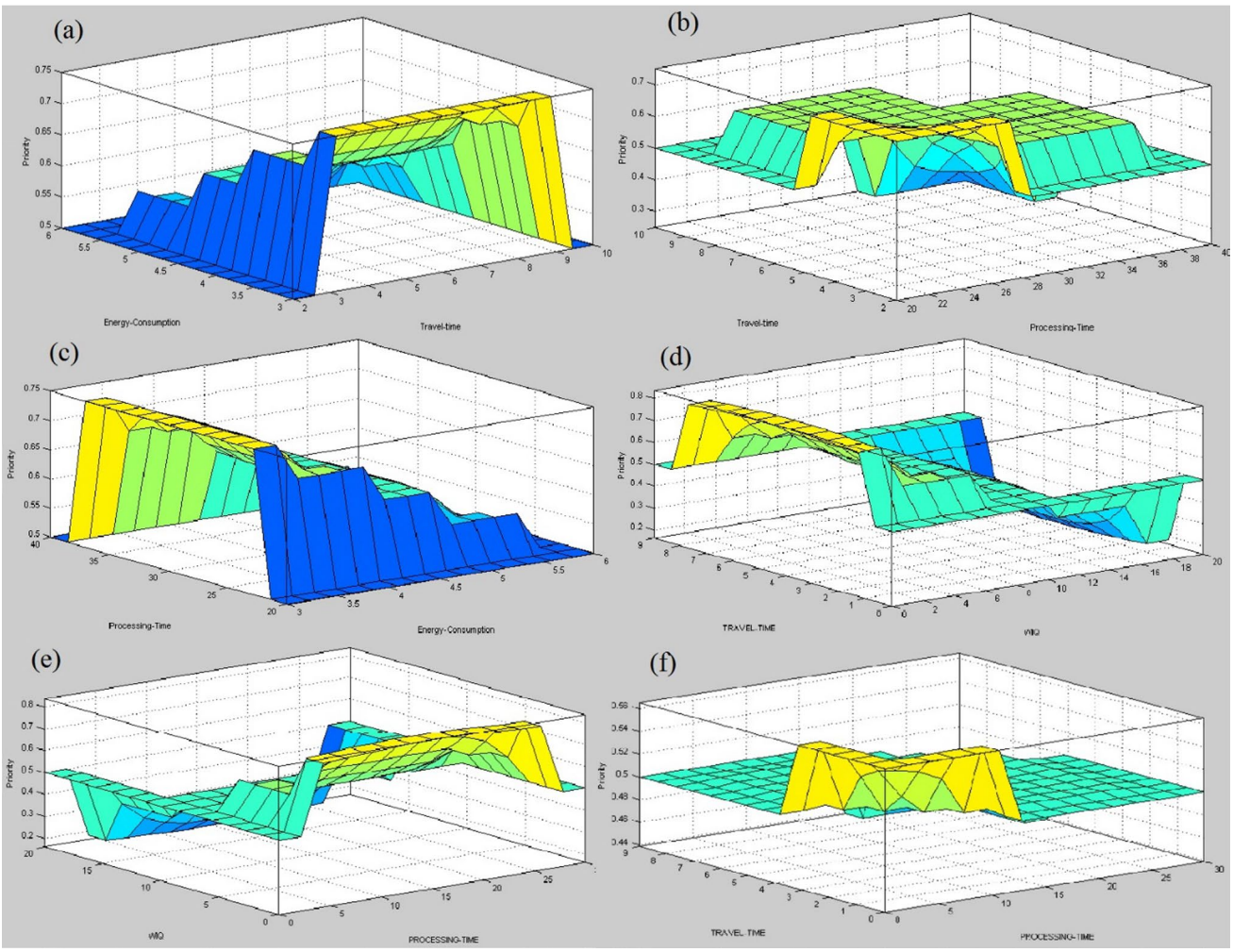


Fig. 4 **a** Impact of energy consumption and travel time on job prioritization, **b** Impact of travel time and processing time on job prioritization, **c** Impact of processing time and profit over cost on job prioritization, **d** impact of travel time and WIQ on crankcase cover route selection, **e** Impact of processing time and WIQ on crankcase cover route selection, **f** Impact of processing time and travel time on crankcase cover route selection

Table 8 Crankcase cover priority using fuzzy reasoning tool

Route (M/C sequence)	KWPG	K38	206G	KTE
DCM-MF-DM-VMC-BM-PS	0.585	0.625	0.553	0.516
DCM-CF-DM-VMC-BM-PS	0.514	0.563	0.5	0.469
DCM-VMC-DM-VMC-BM-PS	0.553	0.606	0.48	0.165

Table 9 Crankcase cover route selection using the fuzzy set approach

Crankcase cover	$u_j(i)_{\max}$
KWPG	0.657
K38	0.619
206G	0.695
KTE	0.738

dynamic uncertainty. This would involve optimizing the schedule with consideration to uncertain and variable processing times, which could improve the overall efficiency of the manufacturing process.

- In addition, future research could explore the integration of machine learning techniques with fuzzy reasoning tool to improve the accuracy and efficiency of the scheduling process. This could involve the use of neural networks or other forms of artificial intelligence to better handle uncertainty and variability in the manufacturing process. Overall, the use of fuzzy reasoning tool and other advanced techniques in manufacturing scheduling has the potential to greatly improve efficiency and reduce energy consumption in a wide range of industries.

Table 10 Final sequence obtained by the fuzzy logic system and fuzzy set approach

The final sequence obtained by the fuzzy reasoning tool		The final sequence obtained by the fuzzy set approach	
Crankcase cover	Optimum route (M/C sequence)	Crankcase cover	Optimum route (M/C sequence)
KWPG	DCM-MF-DM-VMC-BM-PS	KWPG	**DCM-MF-DM-VMC-BM-PS** DCM-CF-DM-VMC-BM-PS
K38	DCM-MF-DM-VMC-BM-PS	K38	**DCM-MF-DM-VMC-BM-PS** DCM-CF-DM-VMC-BM-PS
206G	DCM-MF-DM-VMC-BM-PS	206G	**DCM-MF-DM-VMC-BM-PS** DCM-CF-DM-VMC-BM-PS
KTE	DCM-MF-DM-VMC-BM-PS	KTE	**DCM-MF-DM-VMC-BM-PS**

Author Contributions Both authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by author 1. The first draft of the manuscript was written by author 1, and author 2 revises the manuscript and validate the results. Both authors read and approved the final manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

1. G. Mouzon, M.B. Yildirim, J. Twomey, Operational methods for minimization of energy consumption of manufacturing equipment. Int. J. Prod. Res. **45**(18–19), 4247–4271 (2007)
2. D. Tang, M. Dai, M.A. Salido, A. Giret, Energy-efficient dynamic scheduling for a flexible flow shop using an improved particle swarm optimization. Comput. Ind. **81**, 82–95 (2016)
3. M. Mori, M. Fujishima, Y. Inamasu, Y. Oda, A study on energy efficiency improvement for machine tools. CIRP Ann.-Manuf. Technol. **60**(1), 145–148 (2011)
4. S. Kara, S. Manmek, C. Herrmann, Global manufacturing and the embodied energy of products. CIRP Ann.-Manuf. Technol. **59**(1), 29–32 (2010)
5. C. Herrmann, S. Thiede, S. Kara, J. Hesselbach, Energy oriented simulation of manufacturing systems-concept and application. CIRP Ann. - Manuf. Technol. **60**(1), 45–48 (2011)
6. J. Bonvoisin, S. Thiede, D. Brissaud, C. Herrmann, An implemented framework to estimate manufacturing-related energy consumption in product design. Int. J. Comput. Integr. Manuf. **26**(9), 866–880 (2013)
7. X. Wang, Y. Dai, Z. Jia, Energy-efficient on/off control in serial production lines with Bernoulli machines. Flex. Serv. Manuf. J. (2022). https://doi.org/10.1007/s10696-022-09481-8
8. E.A. Abdelaziz, R. Saidur, S. Mekhilef, A review on energy saving strategies in industrial sector. Renew. Sustain. Energy Rev. **15**(1), 150–168 (2011)
9. C. Gahm, F. Denz, M. Dirr, A. Tuma, Energy-efficient scheduling in manufacturing companies: a review and research framework. Eur. J. Oper. Res. **248**(3), 744–757 (2016)
10. Y. Li, Y. He, Y. Wang, P. Yan, X. Liu, A framework for characterising energy consumption of machining manufacturing systems. Int. J. Prod. Res. **52**(2), 314–325 (2014)
11. J. Wang, F. Qiao, F. Zhao, J.W. Sutherland, Batch scheduling for minimal energy consumption and tardiness under uncertainties: a heat treatment application. CIRP Ann. Manuf. Technol. **65**(1), 17–20 (2016)
12. S.A. Mansouri, E. Aktas, U. Besikci, Green scheduling of a two-machine flowshop: trade-off between makespan and energy consumption. Eur. J. Oper. Res. **000**, 1–17 (2015)
13. M.M. Aghelinejad, Y. Ouazene, A. Yalaoui, Production scheduling optimisation with machine state and time-dependent energy costs. Int. J. Prod. Res. **56**(16), 5558–5575 (2018)
14. J. Yan, L. Li, F. Zhao, F. Zhang, Q. Zhao, A multi-level optimization approach for energy-efficient flexible flow shop scheduling. J. Clean. Prod. **137**, 1543–1552 (2016)
15. G. May, B. Stahl, M. Taisch, V. Prabhu, Multi-objective genetic algorithm for energy-efficient job shop scheduling. Int. J. Prod. Res. **53**(23), 7071–7089 (2015)
16. H. Nanvala, Use of fuzzy logic approaches in scheduling of FMS: a review. Int. J. Comput. Sci. Eng. **3**(4), 1734–1739 (2011)
17. K.T. Fang, B.M.T. Lin, Parallel-machine scheduling to minimize tardiness penalty and power cost. Comput. Ind. Eng. **64**(1), 224–234 (2013)
18. C.A. Garcia-Santiago, J. Del Ser, C. Upton, F. Quilligan, S. Gil-Lopez, S. Salcedo-Sanz, A random-key encoded harmony search approach for energy-efficient production scheduling with shared resources. Eng. Optim. **47**(11), 1481–1496 (2015)
19. M.H. Fazel Zarandi, F. Dorry, A hybrid fuzzy PSO algorithm for solving steelmaking-continuous casting scheduling problem. Int. J. Fuzzy Syst. **20**, 219–235 (2018)
20. C. Lu, L. Gao, X. Li, Q. Pan, Q. Wang, Energy-efficient permutation flow shop scheduling problem using a hybrid multi-objective backtracking search algorithm. J. Clean. Prod. **144**, 228–238 (2017)
21. Y. Liu, N. Lohse, S. Petrovic, N. Gindy, An investigation into minimising total energy consumption, total energy cost and total tardiness based on a rolling blackout policy in a job shop. IFIP Adv. Inf. Commun. Technol. **397**, 103–110 (2013)
22. H. Mokhtari, A. Hasani, An energy-efficient multi-objective optimization for flexible job-shop scheduling problem. Comput. Chem. Eng. **104**, 339–352 (2017)
23. Z. Jiang, L. Zuo, E. Mingcheng, Study on multi-objective flexible job-shop scheduling problem considering energy consumption. J. Ind. Eng. Manag. **7**(3), 589–604 (2014)
24. X. Wu, Y. Sun, A green scheduling algorithm for flexible job shop with energy-saving measures. J. Clean. Prod. **172**, 3249–3264 (2018)

25. X. Gong, T. De Pessemier, W. Joseph, L. Martens, An energy-cost-aware scheduling methodology for sustainable manufacturing. Proc. CIRP **29**, 185–190 (2015)
26. A.A.G. Bruzzone, D. Anghinolfi, M. Paolucci, F. Tonelli, Energy-aware scheduling for improving manufacturing process sustainability: a mathematical model for flexible flow shops. CIRP Ann. - Manuf. Technol. **61**(1), 459–462 (2012)
27. L. Li, H. Huang, F. Zhao, Z. Liu, Operation scheduling of multi-hydraulic press system for energy consumption reduction. J. Clean. Prod. **165**, 1407–1419 (2017)
28. G. Chen, L. Zhang, J. Arinez, S. Biller, Energy-efficient production systems through schedule-based operations. IEEE Trans. Autom. Sci. Eng. **10**(1), 27–37 (2013)
29. Y.C. Choi, Dispatching rule-based scheduling algorithms in a single machine with sequence-dependent setup times and energy requirements. Proc. CIRP **41**, 135–140 (2016)
30. S. Lee, B. Do Chung, H.W. Jeon, J. Chang, A dynamic control approach for energy-efficient production scheduling on a single machine under time-varying electricity pricing. J. Clean. Prod. **165**, 552–563 (2017)
31. L. Merkert, I. Harjunkoski, A. Isaksson, S. Säynevirta, A. Saarela, G. Sand, Scheduling and energy Industrial challenges and opportunities. Comput. Chem. Eng. **72**, 183–198 (2015)
32. J.R. Duflou et al., Towards energy and resource efficient manufacturing: a processes and systems approach. CIRP Ann. - Manuf. Technol. **61**(2), 587–609 (2012)
33. A. Trianni, E. Cagno, P. Thollander, S. Backlund, Barriers to industrial energy efficiency in foundries: a European comparison. J. Clean. Prod. **40**, 161–176 (2013)
34. T.S. Su, Y.F. Lin, Fuzzy multi-objective procurement/production planning decision problems for recoverable manufacturing systems. J. Manuf. Syst. **1**(37), 396–408 (2015)
35. Q. Zhang, J. Zhou, K. Wang, A.A. Pantelous, An effective solution approach to fuzzy programming with application to project scheduling. Int. J. Fuzzy Syst. **20**, 2383–2398 (2018)
36. R. Wang, J. Zhou, X. Yi, A.A. Pantelous, Solving the green-fuzzy vehicle routing problem using a revised hybrid intelligent algorithm. J. Ambient Intell. Human. Comput. **29**(10), 321–332 (2019)
37. H. Babaei, J. Karimpour, A. Hadidi, Applying hybrid fuzzy multi-criteria decision-making approach to find the best ranking for the soft constraint weights of lecturers in UCTP. Int. J. Fuzzy Syst. **20**, 62–77 (2018)
38. H. Jafari, S. Bateni, P. Daneshvar, S. Bateni, H. Mahdioun, Fuzzy mathematical modeling approach for the nurse scheduling problem: a case study. Int. J. Fuzzy Syst. **18**, 320–332 (2016)
39. E. Trillas, Lotfi A. Zadeh: ON the man and his work. Sci. Iran. **18**(3), 574–579 (2011)
40. A. Tighe, F.S. Smith, G. Lyons, Optimisation enhancement using self-organising fuzzy control. Kybernetes **34**(9–10), 1440–1455 (2005)
41. F.T.S. Chan, H.K. Chan, A. Kazerooni, Real time fuzzy scheduling rules in FMS. J. Intell. Manuf. **14**(3–4), 341–350 (2003)
42. H. Singh, M.M. Gupta, T. Meitzler, Z.G. Hou, K.K. Garg, A.M. Solo, L.A. Zadeh, Real-life applications of fuzzy logic. Adv. Fuzzy Syst. Fuzzy Syst. **1**(2013), 3 (2013)
43. S. Hatagar, S.V. Halase, Three Input–one output fuzzy logic control of washing machine. Int. J. Sci. Res. Eng. Technol. **4**(1), 57–62 (2015)
44. A. Talon, C. Curt, Selection of appropriate defuzzification methods: application to the assessment of dam performance. Exp. Syst. Appl. **15**(70), 160–174 (2017)
45. S. Chawla, Developing fuzzy scheduling system for jobs prioritization & route selection of jobs in FMS. Int. J. Sci. Res. **3**(9), 1310–1313 (2014)

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

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

European Economic Letters
ISSN 2323-5233
Vol 14, Issue 1 (2024)
https://doi.org/10.52783/eel.v14i1.1051
http://eelet.org.uk

Enhancing Quality Management in Ethiopia's Higher Education through Industry 4.0 Technologies

Asalf Habtegeorgis*[1], G.C Maheshwari[2], Rajan Yadav[3]

[1]Ph. D. Research Scholar in Management, Delhi School of Management, Delhi Technological University, India
Email ID: asalf2002@gmail.com
[2]Professor in Finance, School of Management, Delhi Technological University, India
Email ID: gcmaheshwari2004@yahoo.com
[3]Professor in Marketing, School of Management, Delhi Technological University, India
Email ID: rajen.yadav@gmail.com

Abstract

This study explores the multifaceted landscape of challenges faced by Higher Education Institutions (HEIs) and investigates the transformative potential of execution challenges and Industry 4.0 technologies in enhancing quality management within Ethiopian HEIs.

The analysis begins by scrutinizing execution challenges—ranging from implementing quality management practices to technological advancements and effective monitoring. Regression analyses affirm a significant positive correlation between addressing these challenges and elevating educational standards. Notably, each challenge demonstrates a robust impact on various facets of educational quality, emphasizing the critical need to address these challenges for improved quality management within HEIs.
Concurrently, the study delves into the integration of Industry 4.0 technologies—examining their role in bridging identified gaps within Quality Management Systems (QMS) in Ethiopian HEIs. Results underscore the transformative potential of these technologies, particularly in faculty development, information technology, and infrastructure enhancement. The findings reveal a clear correlation between the integration of these technologies and substantial improvements in QMS, emphasizing their crucial role in refining educational infrastructures.
The synergy between addressing execution challenges and leveraging Industry 4.0 technologies emerges as a pivotal strategy for HEIs. The positive relationships unveiled in the analyses underscore the need for comprehensive interventions aimed at mitigating execution challenges while harnessing technological advancements. This holistic approach holds promise in not only enhancing educational standards but also fostering efficient and effective quality management systems within Ethiopian HEIs. This study advocates for a strategic framework integrating targeted solutions to address execution challenges alongside the adoption of innovative technologies. Such concerted efforts will not only drive immediate improvements in educational quality but also establish a foundation for continual enhancement and innovation within Ethiopian HEIs. This comprehensive approach presents an opportunity to propel Ethiopian HEIs toward a future characterized by adaptive, technology-driven, and quality-centric educational frameworks, fostering an environment conducive to continual growth and excellence.

Keywords: Industry 4.0, Higher Education, Quality Management, Challenges, Ethiopia, Educational Technology

Introduction

Higher education institutions worldwide encounter multifaceted challenges that intricately affect the quality of education and its management. These challenges encompass a spectrum of influences ranging from political, economic, and socio cultural to execution-related issues, each exerting a distinct impact on the educational landscape (Addis, 2019). In response to these complexities, educational establishments continually strive to

European Economic Letters
ISSN 2323-5233
Vol 14, Issue 1 (2024)
https://doi.org/10.52783/eel.v14i1.1051
http://eelet.org.uk

maintain and enhance the quality of their academic offerings, faculty competence, organizational efficacy, and infrastructure standards ((Addis et al., 2019).

Political challenges often manifest in policy shifts, governmental decisions, and geopolitical events, significantly shaping the operational landscape of educational institutions (Ali & Johl, 2021). Economic instabilities, characterized by global financial imbalances or crises, pose another formidable hurdle affecting higher education resource allocation and developmental initiatives (Ali & Johl, 2022). Moreover, socio-cultural challenges stemming from diverse cultural norms, values, and societal dynamics undeniably influence educational practices and institutional norms (Ergado, 2019). Execution challenges involving conflicts or coordination issues, further compound these complexities, hindering the effective implementation of quality management practices (Hiran & Henten, 2019).

Understanding the impact of these challenges on critical dimensions of higher education—such as the quality of staff and faculty, organizational development, and infrastructure—is pivotal for instituting strategic interventions and fostering improvements. While prior studies acknowledge these challenges' existence, a nuanced exploration of their specific effects on the quality management and assurance framework remains a critical gap in scholarly discourse ((Hiran & Henten, 2019).

This study aims to address this gap by investigating the challenges faced by the HEIs, how execution challenges uplift their standard of education and have better quality management and the role of Industry 4.0 Technologies in Bridging Gaps within Quality Management Systems in Ethiopian HEIs.

Literature Review

The existing body of literature on enhancing quality management in higher education through the utilization of Industry 4.0 technologies indicates a notable increase in worldwide attention towards harnessing advanced technologies to improve educational outcomes (Pehlivan and Cicek, 2021). Similarly, the research conducted by KI-N highlights the significance of incorporating scientifically validated quality management models, such as the ISO model, to prioritize the quality of higher education and establish benchmarks aligned with global standards (Ali and Johl, 2021). The incorporation of Industry 4.0 technologies corresponds with the suggestions mentioned earlier, providing a systematic strategy to tackle the diverse obstacles identified in Ethiopia's higher education domain.

Socio-cultural challenges, deeply rooted in diverse societal norms and values, significantly impact educational practices and institutional norms (Tamrat, 2020). These challenges, ranging from cultural disparities to inclusivity concerns, pose profound implications for the effective functioning of educational institutions. Execution challenges, such as conflicts or coordination deficiencies, emerge as critical impediments to successfully implementing quality management practices in higher education (Pehlivan & Cicek, 2021). These challenges hinder organizational effectiveness, limit innovation, and impede the attainment of quality benchmarks. While studies acknowledge the existence of these challenges, gaps persist in understanding their specific implications for the quality of staff and faculty, organizational development, and infrastructure within higher education. Research efforts have predominantly highlighted the challenges' existence rather than providing comprehensive insights into their intricate impacts on various dimensions of educational quality management (Mhlanga, 2021).

This study seeks to bridge this gap by investigating the nuanced interactions between political, economic, socio-cultural, and execution-related challenges and their effects on the quality management and assurance framework within higher education institutions. By comprehensively synthesizing existing literature, this review sets the stage for the in-depth exploration of these challenges' implications on educational quality, paving the way for informed interventions and policy formulation.

Objective

- To analyse the challenges being faced by the HEIs and how execution challenges uplift their standard of education and have better quality management.

European Economic Letters
ISSN 2323-5233
Vol 14, Issue 1 (2024)
https://doi.org/10.52783/eel.v14i1.1051
http://eelet.org.uk

- To evaluate the Role of Industry 4.0 Technologies in Bridging Gaps within Quality Management Systems in Ethiopian HEIs.

Hypothesis

Ha1: Execution challenges positively impact the standards of education and enhance the quality management practices within HEIs.

Ha2: The integration of Industry 4.0 technologies significantly correlates with addressing and bridging identified gaps within quality management systems in HEIs.

Research Methodology

The research employed a deductive methodology, utilizing quantitative instruments for data gathering, examination, and deduction. Using Yamane's approach, a sample size of 500 people, comprising instructors and students from specific universities, was calculated. Participant surveys were carefully constructed and distributed with the assistance of professional enumerators. The statistical study, which included regression analysis to determine the relationship between quality management indicators and academic results, was made easier using SPSS software. Ethical factors, including informed permission and anonymity, were closely monitored throughout the data-gathering procedure, guaranteeing the study's integrity and adherence to ethical research principles.

Analysis

Objective 1: To analyse the challenges being faced by the HEIs and how execution challenges uplift their standard of education and have better quality management.

Ha1: Execution challenges positively impact the standards of education and enhance the quality management practices within HEIs.

Table 1: Regression Analysis

Challenge	Coefficient	T-statistic	p-value	R^2	Adjusted R^2	F ratio
Constant	0.922	7.249	0.000	0.477	0.472	106.293
ecc1	0.338	8.046	0.000			
ecc2	0.173	2.741	0.006			
ecc6	0.118	2.049	0.041			
ecc7	0.136	2.386	0.017			

The regression analysis depicted in Table 1 demonstrates a significant and positive relationship between execution challenges and the enhancement of quality management within Higher Education Institutions (HEIs). The coefficients associated with execution challenges, including the implementation of quality management practices, information and communication technology, student and faculty participation, and monitoring, evaluation, and reporting, all exhibit positive values. These values signify that as these challenges are addressed or improved upon, there is a tangible upliftment observed in various facets of HEIs, such as staff and faculty quality, organizational development, and infrastructure. Moreover, the statistically significant p-values for each challenge indicate their substantial impact on elevating standards of education and fostering better quality management practices within HEIs. Hence, the findings strongly support the hypothesis, suggesting that effectively tackling execution challenges correlates with an enhancement in educational standards and the overall quality management framework within these institutions.

European Economic Letters
ISSN 2323-5233
Vol 14, Issue 1 (2024)
https://doi.org/10.52783/eel.v14i1.1051
http://eelet.org.uk

The analysis supports the hypothesis (Ha1) indicating a significant positive impact of execution challenges on the quality management framework within Higher Education Institutions (HEIs). These findings, with statistically significant p-values below 0.05, affirm that addressing execution challenges correlates with improvements in educational standards and the overall quality management practices within HEIs.

Objective 2: To evaluate the Role of Industry 4.0 Technologies in Bridging Gaps within Quality Management Systems in Ethiopian HEIs.

Ha2: The integration of Industry 4.0 technologies significantly correlates with addressing and bridging identified gaps within quality management systems in HEIs.

Table 2: Regression Analysis

Quality	Coefficient	T-statistic	p-value	R^2	Adjusted R^2	F ratio
Constant	0.560	4.105	0.000	0.493	0.487	90.580
qi1	0.112	2.198	0.028			
qi2	0.050	0.861	0.390			
qi3	0.191	3.158	0.002			
qi6	0.244	5.278	0.000			
qi7	0.236	5.254	0.000			

The regression analysis in Table 2 outlines the coefficients of Industry 4.0 technologies (qi1, qi2, qi3, qi6, qi7) and their impact on addressing gaps within Quality Management Systems (QMS) in Ethiopian Higher Education Institutions (HEIs). The positive coefficients associated with each technology dimension signify their positive relationship with QMS improvements. Notably, variables like faculty (qi1), information technology (qi3), laboratories (qi6), and playgrounds (qi7) demonstrate substantial impacts on QMS, supported by their statistically significant coefficients and low p-values (all below 0.05). These findings suggest that the integration and enhancement of Industry 4.0 technologies, especially in faculty development, technological infrastructure, and facilities, significantly correlate with bridging identified gaps within QMS in Ethiopian HEIs. Therefore, the results strongly support the hypothesis (Ha2), affirming the crucial role of Industry 4.0 technologies in advancing and refining QMS within HEIs in Ethiopia.

The regression analysis strongly supports the hypothesis (Ha2), indicating a significant positive correlation between Industry 4.0 technology integration and bridging gaps within Quality Management Systems (QMS) in Ethiopian Higher Education Institutions (HEIs). Specifically, dimensions such as faculty development, information technology, laboratories, and playgrounds exhibit substantial coefficients and statistically significant p-values (all below 0.05), underscoring their pivotal role in addressing identified QMS gaps. These findings affirm the crucial contribution of Industry 4.0 technologies to enhancing and refining QMS within Ethiopian HEIs, validating their significance in fostering improvements across diverse facets of the educational infrastructure.

Discussion

The regression analyses conducted on both execution challenges and Industry 4.0 technology integration provide compelling insights into the intricate dynamics of quality management within Higher Education Institutions (HEIs). The results affirm the pivotal role of execution challenges in elevating educational standards and enhancing quality management practices within HEIs, aligning with the hypothesis (Ha1). Notably, factors like implementing quality management practices, improving information technology, fostering student-faculty participation, and effective monitoring and reporting exhibit statistically significant positive associations with various dimensions of

European Economic Letters
ISSN 2323-5233
Vol 14, Issue 1 (2024)
https://doi.org/10.52783/eel.v14i1.1051
http://eelet.org.uk

educational quality, including staff and faculty quality, organizational development, and infrastructure. These findings underscore the critical importance of addressing execution challenges to foster a more robust and efficient quality management framework within HEIs.

Similarly, the analysis regarding Industry 4.0 technologies aligns with the hypothesis (Ha2), revealing a significant positive correlation between these technologies' integration and the bridging of identified gaps within Quality Management Systems (QMS) in Ethiopian HEIs. Specifically, the dimensions of faculty development, information technology, laboratories, and playgrounds demonstrate substantial impacts on QMS improvements, supported by their significant coefficients and low p-values. These findings highlight the transformative potential of Industry 4.0 technologies in addressing deficiencies within QMS, particularly in enhancing faculty competence, technological infrastructure, and institutional facilities.

Overall, these results collectively emphasize the critical interplay between addressing execution challenges and leveraging advanced technological integrations to uplift the educational standards and refine the quality management frameworks within HEIs. The positive relationships unveiled in the analyses underscore the need for strategic interventions aimed at mitigating execution challenges while harnessing the transformative capabilities of Industry 4.0 technologies. This holistic approach holds promise in not only enhancing educational standards but also fostering more efficient and effective quality management systems within Ethiopian HEIs, thereby contributing to the overall advancement and competitiveness of the educational landscape.

Conclusion

The exploration into Industry 4.0 technologies and their integration within Ethiopian Higher Education Institutes (HEIs) unveils a promising trajectory toward redefining quality management practices. The study focused on the execution challenges and the integration of Industry 4.0 technologies within Higher Education Institutions (HEIs) offers profound implications for advancing quality management and educational standards. The findings highlight the intrinsic relationship between addressing execution challenges and leveraging technological advancements to bolster the quality of education within HEIs, underscoring key pathways for transformative improvements.

The robust regression analyses affirm that addressing execution challenges—such as implementing quality management practices, enhancing information technology, fostering student-faculty participation, and ensuring effective monitoring and reporting—positively impacts educational quality. These findings underscore the significance of proactive strategies to mitigate these challenges to promote a more robust and efficient quality management framework within HEIs.

Simultaneously, the pivotal role of Industry 4.0 technologies in bridging identified gaps within Quality Management Systems (QMS) stands evident. Notably, the dimensions of faculty development, information technology, laboratories, and playgrounds exhibit substantial impacts on QMS improvements, elucidating the transformative potential of these technologies in enhancing faculty competence, technological infrastructure, and institutional facilities.

The confluence of these results advocates for a comprehensive approach integrating effective strategies to tackle execution challenges while harnessing the capabilities of advanced technologies. By strategically addressing execution challenges and leveraging Industry 4.0 technologies, HEIs can elevate educational standards, refine quality management practices, and foster an environment conducive to holistic growth and competitiveness.

Moving forward, targeted interventions addressing execution challenges and the strategic adoption of Industry 4.0 technologies should be integral components of HEI policies and initiatives. Such concerted efforts will drive immediate improvements in educational quality and establish a foundation for continual enhancement and innovation within Ethiopian HEIs. This comprehensive approach holds promise in shaping a future where educational institutions thrive through adaptive, technology-driven, and quality-centric frameworks, fostering an environment conducive to continual growth and excellence in higher education.

European Economic Letters
ISSN 2323-5233
Vol 14, Issue 1 (2024)
https://doi.org/10.52783/eel.v14i1.1051
http://eelet.org.uk

Reference

1. Addis, S. (2019). An exploration of quality management practices in the manufacturing industry of Ethiopia. *The TQM Journal*, 32(1), pp.127–142. doi:https://doi.org/10.1108/tqm-01-2019-0031.
2. Addis, S., Dvivedi, A. and Beshah, B. (2019). Quality management as a tool for job satisfaction improvement in low-level technology organizations: the case of Ethiopia. *Production Planning & Control*, 30(8), pp.665–681. doi:https://doi.org/10.1080/09537287.2019.1574510.
3. Ali, K. and Johl, S.K. (2021). Soft and hard TQM practices: future research agenda for industry 4.0. *Total Quality Management & Business Excellence*, 33(13-14), pp.1625–1655. doi:https://doi.org/10.1080/14783363.2021.1985448.
4. Ali, K. and Johl, S.K. (2022). Critical success factors of total quality management practices using Pareto analysis. *International Journal of Productivity and Quality Management*, 36(3), p.353. doi:https://doi.org/10.1504/ijpqm.2022.124704.
5. Ergado, A.A. (2019). Exploring the Role of Information and Communication Technology for Pedagogical Practices in Higher Education: Case of Ethiopia. *International Journal of Education and Development using Information and Communication Technology*, [online] 15(2), pp.171–181. Available at: https://eric.ed.gov/?id=EJ1220775.
6. Hiran, K.K. and Henten, A. (2019). An integrated TOE–DoI framework for cloud computing adoption in the higher education sector: case study of Sub-Saharan Africa, Ethiopia. *International Journal of System Assurance Engineering and Management*. doi:https://doi.org/10.1007/s13198-019-00872-z.
7. Mhlanga, D. (2021). Artificial Intelligence in the Industry 4.0, and Its Impact on Poverty, Innovation, Infrastructure Development, and the Sustainable Development Goals: Lessons from Emerging Economies? *Sustainability*, [online] 13(11), p.5788. doi:https://doi.org/10.3390/su13115788.
8. Pehlivan, D. and Cicek, K. (2021). A knowledge-based model on quality management system compliance assessment for maritime higher education institutions. *Quality in Higher Education*, pp.1–25. doi:https://doi.org/10.1080/13538322.2021.1905654.
9. Sukkar, M.Y. (2022). Quality management in higher education institutions (HEI). *Khartoum Medical Journal*, 10(3). doi:https://doi.org/10.53332/kmj.v10i3.660.
10. Tamrat, W. (2020). The nuts and bolts of quality assurance in Ethiopian higher education: practices, pitfalls, and prospects. *Journal of Education Policy*, 37(3), pp.443–460. doi:https://doi.org/10.1080/02680939.2020.1852604.

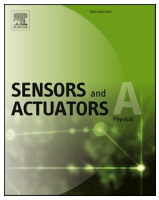
Sensors & Actuators: A. Physical 369 (2024) 115151



Contents lists available at ScienceDirect

Sensors and Actuators: A. Physical

journal homepage: www.journals.elsevier.com/sensors-and-actuators-a-physical



Facile synthesized Sb_2S_3 based high-performance visible photodetector for weak optical signal detection

Pukhraj Prajapat [a,b], Pargam Vashishtha [a,b,c], Deependra Kumar Singh [a], Rajat Kumar Mukherjee [a], Jai Tawale [a], Lalit Goswami [a], Govind Gupta [a,b,*]

[a] CSIR-National Physical Laboratory, Dr K. S. Krishnan Road, New Delhi 110012, India
[b] Academy of Scientific and Innovative Research, CSIR-HRDC Campus, Ghaziabad, Uttar Pradesh 201002, India
[c] School of Electrical and Electronic Engineering, RMIT University, Melbourne, VIC, 3000, Australia

ARTICLE INFO

Keywords:
Sb_2S_3
Photodetector
Li-Fi
Visible Light Communication
Weak Light Detector

ABSTRACT

This report is enlightening a novel approach to fabricating Sb_2S_3-based photodetector that can be used as an efficient receiver with the capability to detect weak optical signals in the visible light communication system. In place of several limitations posed by traditional wireless fidelity networks, this study also provides further evidence supporting the application of the new Light Fidelity technology. The solid-state reaction method was used to synthesize Sb_2S_3 material. A thin film of Sb_2S_3 was then deposited on a SiO_2 substrate via thermal evaporation, and then MSM (metal semiconductor metal) type device was fabricated. The fabricated photodetector demonstrated excellent performance with a responsivity of 2635 mAW^{-1}, detectivity of 10^{11} Jones order, and external quantum efficiency of 615%. This device is remarkable for its ability to detect optical signals with low power, with an estimated detection capability of up to 2.8×10^{-12} $WHz^{-1/2}$ The device also boasts a faster response time, with a rise and decay time of approximately 130 ms. Additionally, it is highly durable and can withstand high voltages, making it an ideal device for visible light communication, even with weak optical signals.

1. Introduction

In recent years, significant advancements have been made in optoelectronics to develop efficient and reliable photodetectors (PDs) for various applications, such as visible light communication (VLC) and more[1]. The universe has expanded to include the endless benefits of wireless communication. Wireless fidelity (Wi-Fi) is one of the most popular wireless communication modalities that connects people worldwide[2]. Nevertheless, this technology has many drawbacks, including external interference, password protection, electromagnetic pollution, and security breaches, among many others [3]. In this context, light fidelity (Li-Fi) enables feasible high-speed data transmission without compromising daily necessities[4]. Li-Fi is a more cost-effective wireless internet system that uses VLC instead of electromagnetic radio waves to attain high-speed internet connections[5]. The main constraints of VLC systems are transmitters, i.e., light-emitting diodes (LEDs), and receivers, i.e., PDs. Li-Fi is gaining interest as a secure, cost-effective, fast communication network that avoids radio frequency constraints and health risks [6]. For the detection of weak optical signals for Li-Fi applications, PDs with high sensitivity, stability, and repeatability with high detectivity (D) and responsivity (R) are crucial because the generated signals lost their signal-to-noise ratio due to traveling medium channel length and other external interferences [6].

PD is a device that generates an electrical current proportional to the optical power incident on its active surface. This fundamental approach enables the development of visible-light communication technology that uses an LED to provide both lighting and wireless communication, known as Li-Fi. As a result, a variety of effective PDs, including Si [7], GaN [8–11], Ga_2O_3 [12], GaAs [13], perovskite materials [14], and other materials [15–21] are currently available. Recently, 2D material-based PDs such as transition metal dichalcogenides (TMDCs), IV-VI group, IV-V group, and transition metal oxides (TMOs) are also in the limelight due to their various essential properties such as quantum confinement, tunable bandgap, layer-by-layer structure, and highly efficient opto-electrical transport [22,23]. Among them, antimony sulfide (Sb_2S_3) is a multilayer chalcogenide semiconductor and a member of the VI-V family. It is well recognized for its numerous applications, such as thermoelectric cooling systems, solar energy conversion, and

* Corresponding author at: CSIR-National Physical Laboratory, Dr K. S. Krishnan Road, New Delhi 110012, India.
E-mail addresses: govind@nplindia.org, govindnpl@gmail.com (G. Gupta).

https://doi.org/10.1016/j.sna.2024.115151
Received 24 November 2023; Received in revised form 26 January 2024; Accepted 5 February 2024
Available online 7 February 2024
0924-4247/© 2024 Elsevier B.V. All rights reserved.

photoconductive targets for VLC, television cameras, etc. [24,25]. In addition to the advantage of 2D materials, Sb_2S_3 is a promising photodetection material due to its high light absorption coefficient ($>10^5$ cm^{-1}), low cost, and nontoxic nature [26,27]. Recent studies reveal that the Sb_2S_3 thin-film PDs are highly stable and ideal for detecting weak signals in the visible light spectrum [28]. As far as Sb_2S_3 growth is concerned, primarily chemical vapor deposition technique has been used, wherein low dimensional materials such as nanowires and nanotubes like nanostructures have been grown [29,30]. Sb_2S_3 thin-film PDs have significant advantages in terms of uniformity, compatibility, and affordability. Therefore, unlike chemical vapor deposition, it is desirable to fabricate large-area, crystalline Sb_2S_3 thin films using the simplest growth method, i.e., thermal evaporation [14].

This study introduces the solid-state method for synthesizing high-quality micro rods of Sb_2S_3. The micro rods were crushed in powder form and employed to deposit the thin film through thermal evaporation. The film quality was analyzed by various characterization techniques, and top electrode metal semiconductor metal deceive was fabricated for the electric measurements. The fabricated PD devices demonstrate a high responsivity (R) of 2635 mAW^{-1} and a detectivity (D) of 6.7×10^{11} Jones at a lower optical power of 10 μW, using an optical excitation wavelength of 532 nm. Moreover, the devices showcase remarkable environmental stability even when subjected to higher applied voltages. This robustness suggests numerous applications for Sb_2S_3 PDs, including their potential as optical signal receivers for Li-Fi text transmission systems. The integration of these PDs into Li-Fi technology holds promise for enhancing data transmission capabilities, capitalizing on their high-performance characteristics and stability even in demanding operational conditions. Overall, this research contributes a solid foundation for the development of efficient and reliable Sb_2S_3-based PDs with implications for advancing optical communication systems.

2. Experimental

To synthesize the Sb_2S_3 material, a mixture of high purity (99.999%) antimony (Sb) and sulfur (S) powder in a stochiometric ratio (from Sigma Aldrich) was used. The mixture was vacuum-sealed in a quartz tube and heated carefully (temp increases by 300 °C per hour) to prevent an explosion due to the high vapor pressure of sulfur. After being maintained at 700 °C for 48 h, the mixture was thoroughly homogenized and cooled gradually at 10 °C per hour. As a result, cracking caused by the melt's thermal expansion during solidification was avoided, and after compaction, micro rods of a dark greyish tint were produced. A thin film of Sb_2S_3 was deposited on SiO_2 by utilizing the thermal evaporation technique at a base pressure of 2.4×10^{-6} torr, using crushed powder of Sb_2S_3 micro rods and a tungsten boat (resistively heated) as the evaporator source. Substrates (SiO_2) were cleaned with acetone, isopropanol, and deionized water before being introduced into the vacuum system, followed by drying by nitrogen. After the deposition of Sb_2S_3 thin-film, the sample was annealed at 300 °C for 2 h in the presence of nitrogen to achieve high crystalline quality. The morphology and structure of the Sb_2S_3 thin films were analyzed by field-emission scanning electron microscopy (FESEM, Hitachi S-4800), X-ray diffraction (XRD) with Cu Kα radiation (Shimadzu XRD-6000), respectively. The surface chemistry of the thin film was characterized using Omicron X-ray spectroscopy with Al k_α Source. The UV-Vis absorbance of the film was measured using a UV-3600 spectrophotometer, and the photoluminescence (PL) was measured using an Edinburgh FLS980 D2D2 system with a 532 nm excitation source (DPSS laser). A 532 nm argon-ion laser was used to perform backscattering-based Raman spectroscopy. Further, the grown Sb_2S_3 thin film was utilized to fabricate visible-PD in MSM device structure by depositing metal electrodes of Au (~ 90 nm)/Ti (~ 10 nm) using the DC sputtering system. A probe station setup (S10 Triax probe station with Keithley 2450 source meter) was used to perform wavelength-correlated bias & power-dependent transient photo response measurements on the competitive advantage by making a 1.18 $\times 10^{-3}$ cm^2 active area. With light output intensities ranging from 10 μW to 356 μW, a 532 nm laser was employed to provide illumination. The dark room served as the setting for all of the photo response tests. Fig. 1.

3. Results and discussion

The structural analysis of the Sb_2S_3 thin film was performed using an XRD technique. Fig. 2(a) shows the XRD pattern, peaks at positions 15.4°, 17.6°, 25.01^0, 35.4°, and 48.3°, corresponding to the (200), (120), (130), (311), and (600) planes of Sb_2S_3 (JCPDF card no. 42-1393) [31]. To further investigate the crystal quality, the full width at half-maximum (FWHM) of the intense peak is we calculated (0.3304 degrees), which indicates strong crystallinity of the film. Moreover, the crystallite size of the thin film is also determined using the below Eq. 1.

$$D = \frac{K\lambda}{\beta \cos\theta} \quad (1)$$

The calculated value of the crystallite size is ~ 43 nm.

Furthermore, Raman spectroscopy is a technique used to analyze material structure by determining crystal orientation and chemical bond vibration symmetry. [32] The Raman spectra shown in Fig. 2(b) exhibit distinct Raman peaks near 310 cm^{-1}, 305 cm^{-1}, 308 cm^{-1}, 299 cm^{-1}, 280 cm^{-1}, 237 cm^{-1}, 188 cm^{-1}, 155 cm^{-1}, 126 cm^{-1}, 97 cm^{-1}, and 71 cm^{-1} which are consistent with those previously reported in the literature.[29,33] The bands at 280 cm-1 and 299 cm-1 are attributed to the symmetric vibrations of Sb_2S_3 pyramidal units with C3v symmetry. Room-temperature photoluminescence (RTPL) spectroscopy was used to analyze the optical properties of the Sb_2S_3 film. The PL spectra in Fig. 2 (c) showed a centered peak at 1.75 eV, which suggested a strong emission of Sb_2S_3. The corresponding luminescence peak positions of the Sb_2S_3 film were around 710 nm, which is in accord with crystalline Sb_2S_3. Further, absorption spectra were used to investigate the optical properties of the synthesized Sb_2S_3 thin film, which exhibited continuous absorption in visible regions, indicating potential applications in solar cells and visible light PDs, as shown in Fig. 2(d). The energy band-gap of the Sb_2S_3 thin film is estimated by the blow formula.

$$(\alpha h\nu)^2 = A(h\nu\text{-}Eg)$$

where α is the absorbance coefficient, h is the Planck constant, A is the characteristic constant of semiconductors, E_g is the optical band gap energy, and ν is the frequency. The estimated value of the band gap of Sb_2S_3 is 1.90 eV, which is in good agreement with the earlier report. [14, 34].

The surface morphology of the Sb_2S_3 thin films that were grown was examined using a FESEM. In Fig. 2(e) & S1, the top view of the Sb_2S_3 film displays some defects due to increased roughness and certain lines from the deposition process. Additionally, morphology was examined by cross-section images at 200 nm resolution (Fig S2) & EDX, which showed the ratio of Sb and S to be approximately 37% and 63%, respectively. Distinct crystalline faces and edges are visible, with a larger grain size suggesting good crystallinity of the Sb_2S_3 thin film. [35, 36] The efficiency of a thin-film device is highly dependent on grain size. Large grains in the absorption layer have significant minority carrier diffusion lengths and built-in electric fields. [37] Large grains result in fewer grain boundaries, which can improve device performance by reducing interfacial recombination. [38,39]. Therefore, it was revealed that the high-quality Sb_2S_3 thin film consisting of large grains can be achieved after post-annealing at 300 °C. Fig. 2(f) displays the Sb_2S_3 crystal structures as angle ball-and-stick representations. Sb atoms are depicted by the red balls, while the blue balls represent S atoms.

Additionally, XPS analysis was used to determine the Sb_2S_3 thin film's composition and chemical purity. The survey spectrum of Sb_2S_3 is shown in Fig. 3(a). It shows clear peaks from Sb 3d, Sb 4d, C 1 s, S 2p,

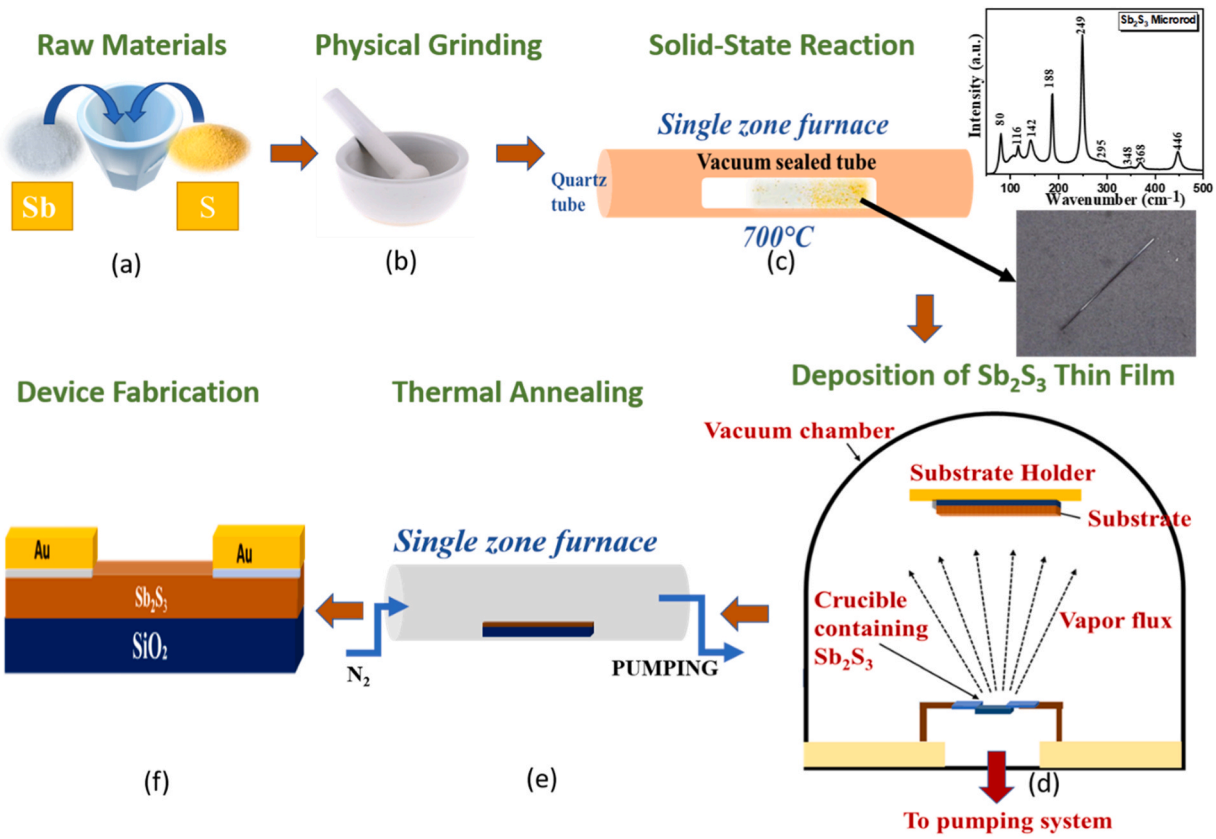


Fig. 1. Schematic diagram of the synthesis process of Sb_2S_3 material and device fabrication process.

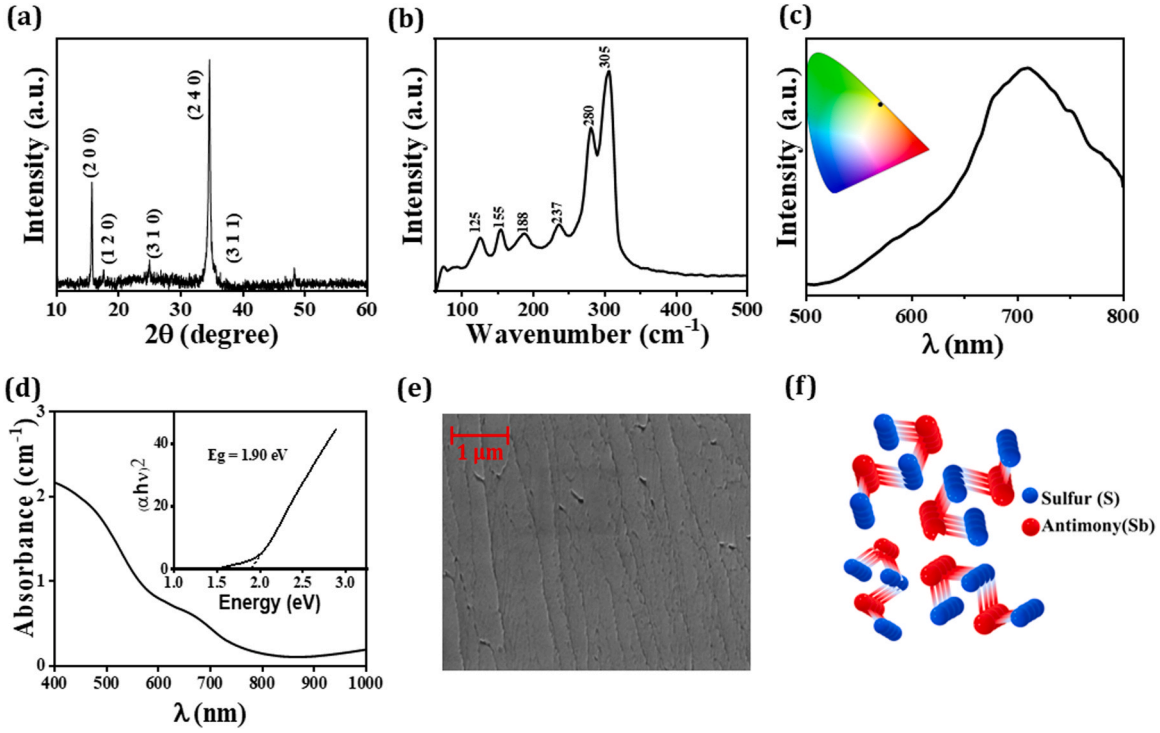


Fig. 2. (a) XRD pattern (b) Raman spectra (c) PL emission spectra (d) Absorption spectra (e) FESEM image (top view) of the Sb_2S_3 film (f) atomic structure of Sb_2S_3.

and S 2 s, along with MNN(Sb) and LMM(S) without any impurities. However, carbon contamination is almost impossible to avoid in most preparations. In order to determine peak binding energy and charge correction precisely, the C1s spectral line at 284.6 eV was used for calibration. The Sb 3d core level spectra of Sb compounds are shown in Fig. 3(b), where the peaks due to Sb-S bond can be observed. The

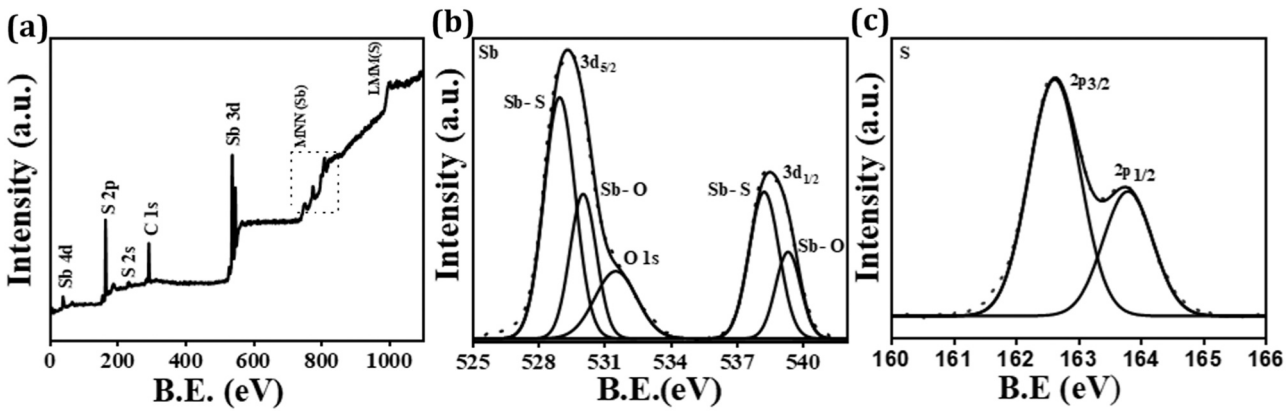


Fig. 3. XPS spectra of the antimony sulfide thin film (a) Survey scan spectra (b) Sb 3d core level spectra (c) S 2p core level spectra.

corresponding Sb $3d_{5/2}$ and Sb $3d_{3/2}$ binding energies are 528.95 eV and 538.25 eV, respectively. The spin-orbit splitting for this material was found to be 9.3 eV, which is consistent with previous reports [40]. It was also revealed that the film oxidized at the surface, and the Sb-O bond was formed, as evident from the peaks Sb $3d_{5/2}$ and Sb $3d_{3/2}$ at 530 eV and 539.3 eV, respectively. Another peak at 531.6 eV was found due to unreacted oxygen on the surface when the sample was exposed to the environment. The investigation involved analyzing the S 2p core level spectra through peak fitting, which identified states at 162.65 eV and 163.8 eV for $S2p_{3/2}$ and $S2p_{1/2}$, respectively. [41] These states are characteristic of the sulfide state. The energy separation of 1.15 eV is consistent with previously reported sulfide values, confirming the chemical state's precise determination.

The photodetection measurements of Sb_2S_3-based PD were carried out under low light illumination to determine its detecting capability in visible light communication applications such as Li-Fi. The impact of wavelength-correlated light illumination of the device, i.e., spectral response, is shown in Fig. 4(a), which displays the change in photocurrent (I_{ph}) with wavelength (λ). The spectral response exhibits maximum current in the visible range for a fixed power. Although the bandgap of Sb_2S_3 is 1.9 eV, the device showed sensitivity to shorter and higher wavelengths, which could be due to several possible factors, such as surface states, defect levels, or other material characteristics. After that, photodetection measurements of the fabricated device were performed using a 532 nm illumination source at different powers, and bias voltages were applied because they align with human visual comfort and enhance user experience, which is designed for VLC applications. The Sb_2S_3 device exhibits photovoltaic behavior, as evidenced by the I-V characteristic of the device with varied power in Fig. 4(b). The illumination power varies from 10 μW to 356 μW, and the maximum estimated I_{ph} is ~287 μA. The channel conductance increases nonlinearly as the bias voltage increases due to the augmentation of photogenerated carriers at higher power density levels. Moreover, the observed asymmetry in the I-V curves may arise from factors such as contact barrier height asymmetry or uneven charge carrier distribution. It is also possible that electrodes have non-uniformity.

The transit photo response (I-T) measurements were performed under laser illumination for repeated on/off cycles spanning 20 s each. Fig. 5(a) displays the bias-dependent properties of the fabricated Sb_2S_3 PD, keeping the optical power fixed at 356 μW. The value of I_{ph} increases with applied bias from 0.5 V to 8 V as the electric field around the conductive channel increases because more photogenerated charge carriers are trapped due to increased drift velocity. Fig. 5(a) & S4 also exhibits that the fabricated PD responds differently to different biases. It can be concluded that Sb_2S_3-based PD exhibits high stability and sensitivity to applied voltages.

The key parameters for the fabricated PD, including R, external quantum efficiency (EQE), noise equivalent power (NEP), and detection (D), were calculated using Eqs. (2–5).

$$R = \frac{I_{ph}}{P_d \times A_o} A \Big/ W \tag{2}$$

Here, A_O and P_d represent the active area of the device (1.18×10^{-3} cm^2) and incident power density, respectively. The EQE is defined as the

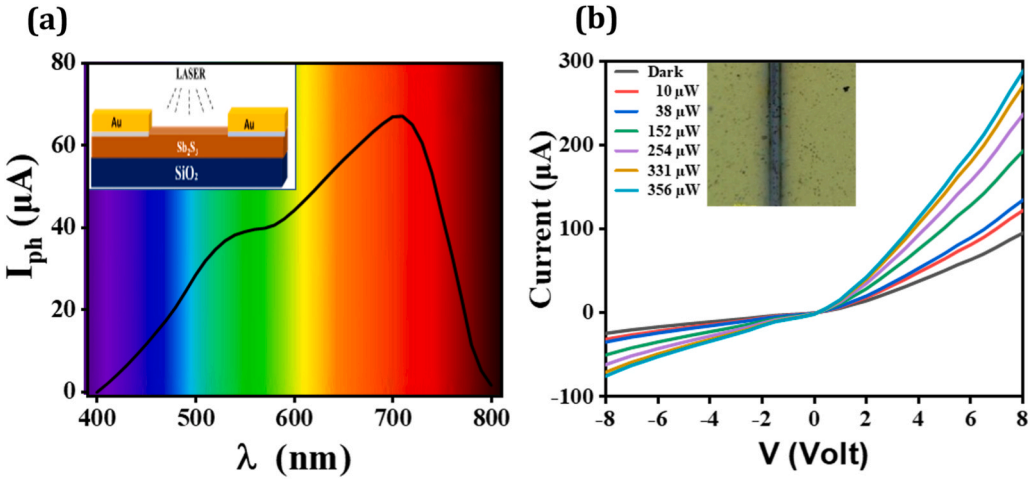


Fig. 4. (a) Spectral response of Sb_2S_3 / SiO_2 devices, inset display schematic device structure (b) Power-dependent I-V characteristics of the Sb_2S_3 / SiO_2 PD under dark and light using 532 nm laser inset shows optical image of the actual device.

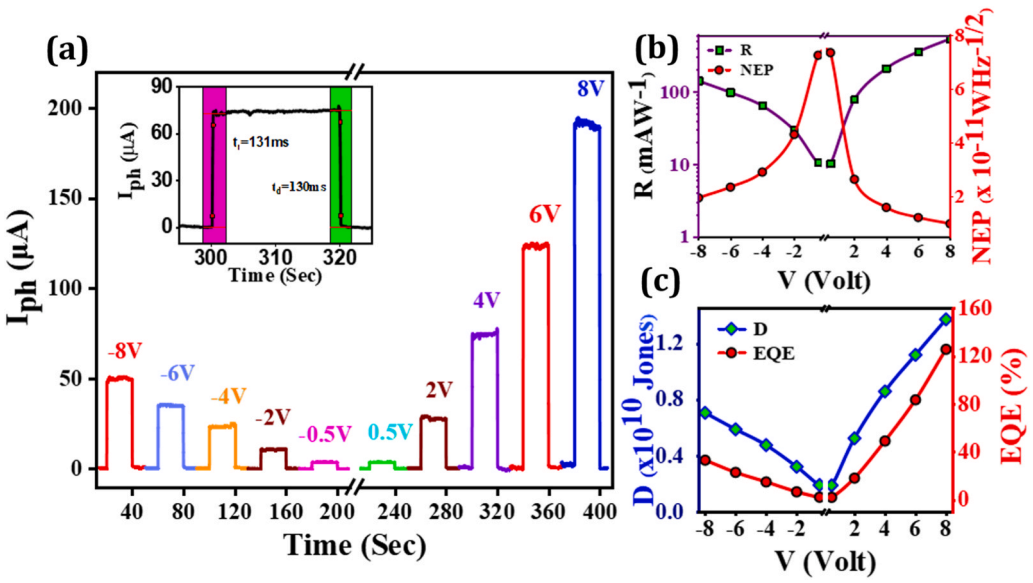


Fig. 5. (a) Time correlated photo response of the Sb_2S_3 PD as a function of applied bias at 356 μW optical power (532 nm); inset display rise time and decay time of a cycle. (b,c) Critical parameters of PDs R, NEP, D, and EQE as a function of applied bias and fixed optical power (356 μW).

photo-conversion efficiency.

$$\mathrm{EQE} = \frac{h \times c \times R}{e \times \lambda} \tag{3}$$

where h, e, c & λ are the Plank constant ($6.62607015 \times 10^{-34}$ m^2 kg / s), electron charge ($1.60217663 \times 10^{-19}$ coulombs), light speed (3×10^8 m/s), and the illumination source's wavelength(532 nm). Furthermore, NEP is the least detectable power density for the dark noise by a PD with a 1 Hz bandwidth frequency and can be defined as

$$\mathrm{NEP} = \frac{\sqrt{2e \times I_d}}{R} \mathrm{WHz}^{-1/2} \tag{4}$$

D is the inverse of the NEP signal.

$$\mathrm{D} = \frac{R\sqrt{A_o}}{NEP} \mathrm{Jones} \tag{5}$$

Fig. 5(b) shows a bias-dependent plot of R and NEP for 532 nm illuminations. The maximum values for R and NEP were determined to be 545 mAW^{-1} and 3.6×10^{-10} $WHz^{-1/2}$, respectively. Other important

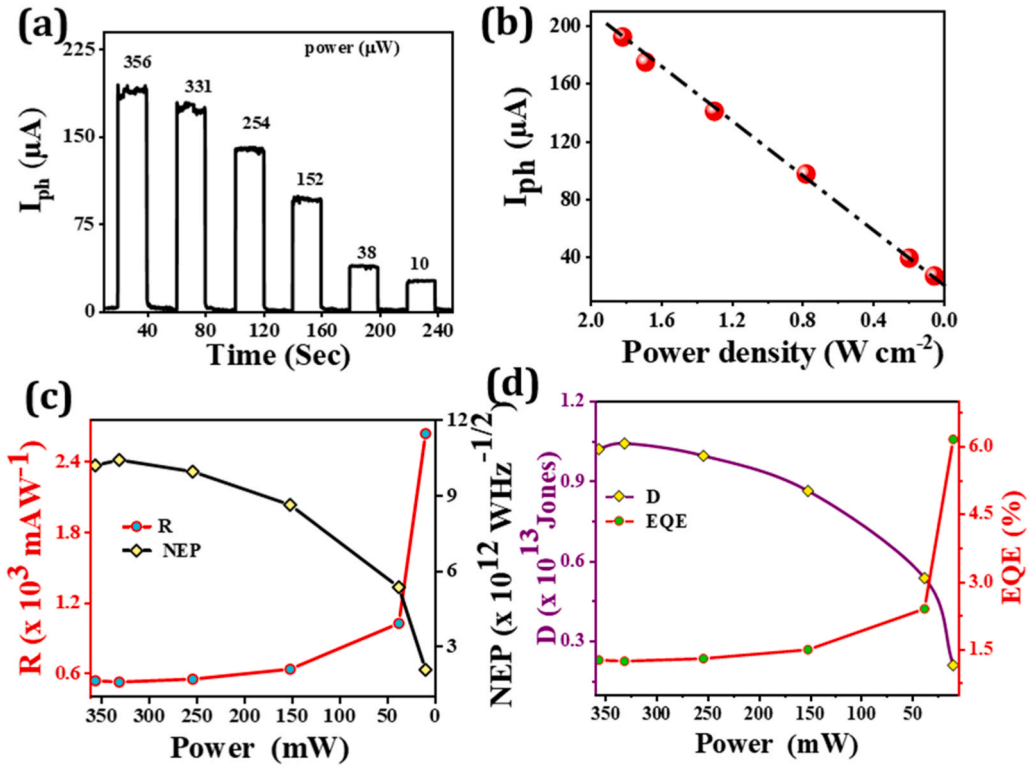


Fig. 6. (a) Photo response of the Sb_2S_3 PD as a function of applied optical power (10 μW to 356 μW) at fixed bias using 532 nm laser (b) I_{ph} vs power density curve (c, d) R, NEP, D, EQE as a function of applied power and fixed bias (8 V) of Sb_2S_3 PD.

parameters, D and EQE, were also displayed in Fig. 5(c). It shows a bias-dependent plot of D and EQE, where a minimum value of D was observed as 1.9×10^8 jones, which increases to 1.37×10^{10} jones on increasing applied bias from 0.5 V to 8 V. The enhancement in the D is due to the increased photo-to-dark current ratio. A bias-dependent plot of EQE is also shown in Fig. 5(c). The EQE value increases as bias increases from 0.5 V to 8 V bias and reaches 123%. EQE follows a similar trend as R and reaches a maximum value at 8 V applied voltage. In common, the EQE value of the PD is 100%, indicating that a single photon has generated a charge carrier that contributed to the current. However, in the fabricated device, the EQE value is more than 100%, and the high EQE values display high internal gain in the fabricated device. Additionally, the response speed (The response speed refers to the duration from 10% to 90% of the highest current) of the device was examined through determining the rise and decay times of the fabricated device, the photocurrent fitting as shown in the inset of Fig. 5(a). The fabricated PD exhibits a fast time-correlated transient response, with rise and decay times as low as ~130 ms.

Fig. 6(a) shows the power-dependent analysis of Sb_2S_3-based device in the visible region where a correlation between incident power intensity and current is observed. For each ON/OFF cycle, the device shows a stable current from 10 μW to 356 μW. As shown in Fig. 6(a), a maximum photocurrent Iph value of 192 μA was demonstrated at 356 μW while a minimum value of 26 μA for 10 μW, which depicts the I-T to visible light. It can be understood that as more photons interact with the material, more electrons are excited to the conduction band, resulting in a high current. The analysis revealed that the fabricated device displays a high response in terms of I_{ph} corresponding to illumination because the photogenerated electron-hole pair would more efficiently move on Sb_2S_3 film. The photocurrent is described as a function of incident illuminance by integrating it with the expression $I \alpha P^{\theta}$. The value of θ is about 0.35, showing the sublinear behavior of PD in Fig. 6(b). The R of the device is calculated with the help of I_{ph} shown in Fig. 6(c). The determined R value is 2.635 AW^{-1} for 10 μW power. The value of R increases exponentially on decreasing incident optical power from 356 μW to 10 μW. Higher R values at low optical power are attributed to minority carrier traps, i.e., hole traps, in the materials when excited by weak light. When a hole enters a hole trap, it cannot recombine with an electron directly; instead, it must be excited into the valence band and recombined. This process is equivalent to increasing the lifetime of recombination of holes, resulting in high R at low incident light. [42].

Fig. 6(c) also displays that the NEP parameter of the fabricated device is as low as 2.8×10^{-12} $WHz^{-1/2}$, respectively. Fig. 6(d) shows the D and EQE performance metrics. It displays a D and EQE power-dependent plot, where a minimum value of D was determined to be 1.38×10^{11} Jones, increasing to 6.7×10^{11} Jones as applied power increased from 10 μW to 356 μW. The computed D value of the fabricated device is encouraging compared to earlier reports, even while operating with a weak signal. The high D value is particularly useful for Li-Fi applications, which require a highly detectable device at the receiver end. A power-dependent plot of EQE (shown in Fig. 6(d)) EQE value increases as power decreases from 356 μW to 10 μW and reaches 615% maximum value at 10 μW power as similar trend of R. A consistent range of R spanning various light intensities is crucial in practical PD applications to derive the precise intensity of received light from the related photocurrent The linear dynamic range (LDR), which is characterized by an optical power margin where the output photocurrent is linearly proportional to the optical signal input, is a constant feature of the linear weak-light response range. The LDR of the fabricated PD was determined to be ~31. A sufficient LDR is a need for weak light sensing since it ensures steady responsiveness from high light to weak light circumstances in Li-Fi technology.

Fig. 7(a) describes the energy level diagrams of the Sb_2S_3-based device with Au/Ti electrodes in isolated conditions. The work function of Sb_2S_3, Ti, and Au are 3.75 eV, 4.3 eV, and 5.1 eV, respectively. Due to the larger work function of Sb_2S_3 (p-type semiconductor) than Ti and Au metal electrodes, the electrons will transfer from metal (low work function) to semiconductor (higher work function) and cause the upward movement of fermi level of semiconductor until the equilibrium condition is reached as shown in Fig. 7(b). If a positive bias (V>0) is applied, then the fermi energy of metal becomes lower than that of

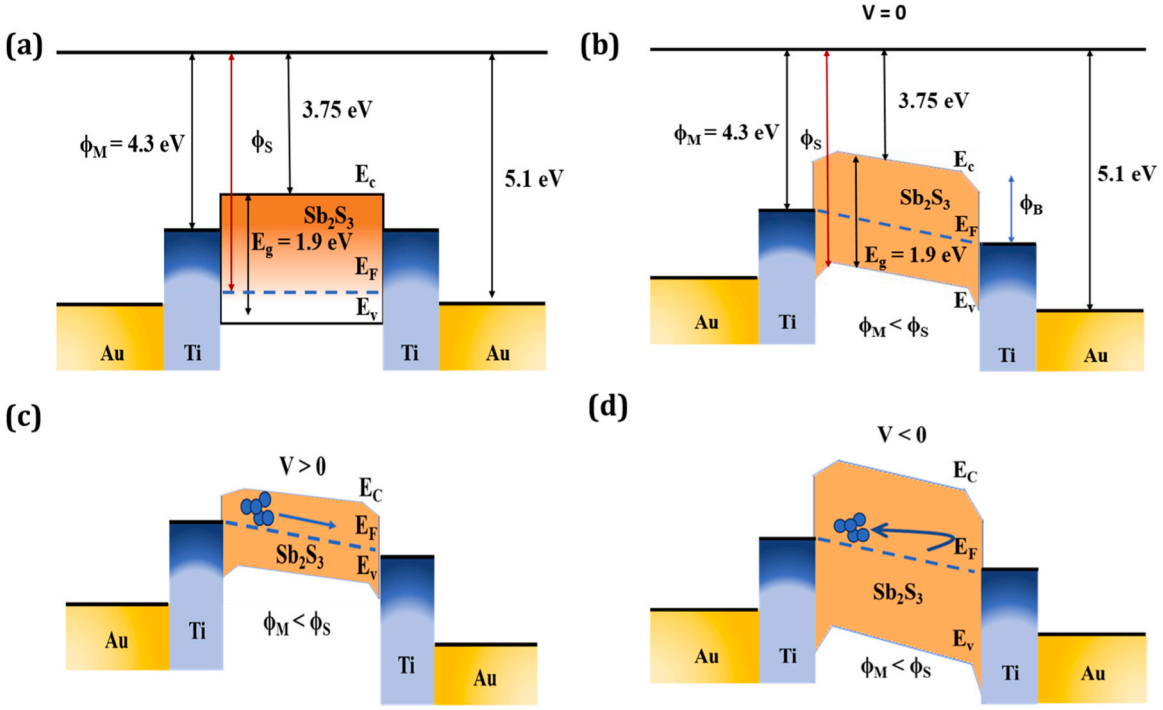


Fig. 7. (a) Position of E_{fm} of Au, Ti, and Sb_2S_3 w.r.t. vacuum level (b, c, d) Band edge alignment of Sb_2S_3 based device at equilibrium (Zero bias), forward bias and reverse bias.

semiconductor, which in turn reduces the barrier height as depicted in Fig. 7(c). In opposite to this, if reverse bias is applied (V<0), then the fermi energy of metal becomes higher than the fermi energy of semiconductor and results in an increment of barrier height shown in Fig. 7 (d). This behavior of MS junction with bias is perfectly in match with the I-V characteristics of the device. Larger bias results in more charge carriers reaching the device's electrodes, which results in a higher photocurrent under forward bias conditions [43].

As shown in Table 1 below, the fabricated device performs better than the previously reported values of Sb_2S_3-based PDs. The thin film is deposited using either spin coating or physical vapor deposition, and all of the reports in Table 1 are examined under almost identical conditions. The Sb_2S_3 thin film has a crystalline structure, which improves device performance as the active material layer. The high crystallinity improves responsiveness with low light-detecting capability and good NEP. Moreover, the Sb_2S_3 thin film-based device's performance can be further improved by adding decorated nanoparticles and heterostructures made of different materials.

Table 1
Sb_2S_3 Based visible photodetector.

S. No	Device	λ (nm)	R (A/W)	D (Jones)	References
1	1D Sb_2S_3	532	0.219	-	[29]
2	FTO/TiO_2/Sb_2S_3/Sb_2Se_3/Au	530	0.340	8.55×10^{12}	[44]
3	Ag/Spiro/Sb_2S_3/FTO(glass)	470	0.087	2.7×10^{9}	[31]
4	Sb_2S_3	-	0.3	1.25×10^{12}	[45]
5	Au-nanoparticles-decorated Sb_2S_3 NW	470	59.5	4.29×10^{10}	[46]
6	Sb_2S_3	625	1.79×10^{-6}	-	[47]
7	PDMS/FTO/TiO_2/Sb_2S_3/Au	530	3.38	2.84×10^{13}	[48]
8	Graphene/Sb_2S_3/TiO_2/FTO	Visible light	0.35	1.13×10^{13}	[49]
9	Sb_2S_3	532	2.635	6.7×10^{11}	This work

4. Conclusion

In summary, high-performance visible light Sb_2S_3-based PDs on silicon oxide substrate were fabricated using a two-step thermal evaporation method. These PD demonstrated remarkable performance in photoconductive visible light, including high stability, excellent repeatability, high responsiveness, and lower cost. PD exhibits enhanced photoconductive behavior, high stability, high D, and weak light detection. The maximum R of 2635 mAW^{-1} at 8 V bias under irradiation of 532 nm was achieved after a thorough investigation of the voltage-dependent and power-dependent on-off switching response and other performance parameters such as D, NEP, EQE, LDR are 6.7×10^{11} jones, 2.8×10^{-12} $WHz^{-1/2}$, 615% and 31.028 respectively. Thus, this study revealed that an Sb_2S_3 device is a good option for high-performing PDs for visible light. It can be helpful for the visible light communication system (Li-Fi) at the receiver edge due to its weak signal detection property and high R. This kind of large-scale Sb_2S_3 is beneficial for commercial application in optical communication.

Author Statement

This report explains a novel method for creating an effective receiver that can detect weak optical signals in a visible light communication system using a Sb_2S_3-based photodetector. High-quality Sb_2S_3 material was synthesized by a novel solid-state approach Sb_2S_3 in the form of micro rods and then thermally evaporated on the SiO_2 substrate for the first time. The fabricated photodetector exhibited outstanding performance as highest responsivity (2635 mAW^{-1}), detectivity (>10^{11} Jones), and external quantum efficiency (615%) at lower optical power of 10 μW with an optical excitation wavelength of 532 nm. Besides, the fabricated device also witnesses the lowest noise equivalent power, which estimates the device detection capability up to 2.8×10^{-12} $WHz^{-1/2}$ order. The fabricated photodetector has also been realized into a fast time-correlated transient response device. Visible light communication works best with this incredibly dependable receiver since it can withstand high voltages and has a weaker optical signal capability.

CRediT authorship contribution statement

Prajapat Pukhraj: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Vashishtha Pargam:** Writing – original draft, Validation, Investigation, Formal analysis. **Singh Deependra K:** Investigation, Formal analysis. **Mukherjee Rajat K:** Investigation. **Tawale Jai:** Investigation. **Goswami Lalit:** Validation, Methodology, Investigation. **Gupta Govind:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Investigation, Funding acquisition, 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

Director, CSIR-NPL, is acknowledged for his continuous support and encouragement. PP, PV is grateful to CSIR & UGC, India, for the financial support under CSIR-JRF & UGC-SRF fellowship, respectively. The authors would like to acknowledge Dr. VPS Awana, Dr. Ajeet Kumar, and Dr. Prathap Pathi for XRD, electrode deposition, and UV-Visible measurements, respectively. D.K.S. is thankful to the SERB National Post-doctoral Fellowship (PDF/2021/002469) for providing financial support.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.sna.2024.115151.

References

[1] Y. Ma, G. Zhou, S. Wang, WiFi sensing with channel state information: a survey, ACM Comput. Surv. 52 (2019) 1–36.
[2] G. Redlarski, B. Lewczuk, A. Zak, A. Koncicki, M. Krawczuk, J. Piechocki, K. Jakubiuk, P. Tojza, J. Jaworski, D. Ambroziak, Ł. Skarbek, D. Gradolewski, The influence of electromagnetic pollution on living organisms: historical trends and forecasting changes, Biomed. Res. Int. 2015 (2015) 1–18.
[3] A.I. Al-Alawi, WiFi technology: future market challenges and opportunities, J. Comput. Sci. 2 (2006) 13–18.
[4] S.V. Kumar, K. Sudhakar, L.S. Rani, Emerging technology Li-Fi over Wi-Fi, Int. J. Invent. Eng. Sci. 2 (2014) 5–6.
[5] H. Haas, L. Yin, Y. Wang, C. Chen, What is LiFi? J. Light. Technol. 34 (2016) 1533–1544.
[6] H. Haas, LiFi is a paradigm-shifting 5G technology, Rev. Phys. 3 (2018) 26–31.
[7] Z. Huang, J.E. Carey, M. Liu, X. Guo, E. Mazur, J.C. Campbell, Microstructured silicon photodetector, Appl. Phys. Lett. 89 (2006) 14–17.
[8] A. Gundimeda, S. Krishna, N. Aggarwal, A. Sharma, N.D. Sharma, K.K. Maurya, S. Husale, G. Gupta, Fabrication of non-polar GaN based highly responsive and fast UV photodetector, Appl. Phys. Lett. 110 (2017).
[9] L. Goswami, N. Aggarwal, P. Vashishtha, S.K. Jain, S. Nirantar, J. Ahmed, M.A. M. Khan, R. Pandey, G. Gupta, Fabrication of GaN nano-towers based self-powered UV photodetector, Sci. Rep. 11 (2021) 1–10.
[10] P. Vashishtha, M. Kumar, P. Prajapat, J. Ahmed, V.N. Singh, G. Gupta, Highly responsive SnSe/GaN heterostructure-based UVC-SWIR broad-band photodetector, Mater. Sci. Semicond. Process. 156 (2023) 107277.

[11] L. Su, Y. Zhang, J. Xie, An all-inorganic $CsPbBr_3$/GaN hetero-structure for a near UV to green band photodetector, J. Mater. Chem. C 10 (2022) 1349–1356.
[12] L. Goswami, N. Aggarwal, R. Verma, S. Bishnoi, S. Husale, R. Pandey, G. Gupta, Graphene quantum dot-sensitized ZnO-nanorod/GaN-nanotower heterostructure-based high-performance UV photodetectors, ACS Appl. Mater. Interfaces 12 (2020) 47038–47047.
[13] W. Liu, M. Wang, C. Xu, S. Chen, X. Fu, One-pot synthesis of ZnO_2/ZnO composite with enhanced photocatalytic performance for organic dye removal, J. Nanosci. Nanotechnol. 13 (2013) 657–665.
[14] L.H. Zeng, S.H. Lin, Z.J. Li, Z.X. Zhang, T.F. Zhang, C. Xie, C.H. Mak, Y. Chai, S. P. Lau, L.B. Luo, Y.H. Tsang, Fast, self-driven, air-stable, and broadband photodetector based on vertically aligned $PtSe_2$/GaAs heterojunction, Adv. Funct. Mater. 28 (2018) 1–11.
[15] X. Yu, H.N. Tsao, Z. Zhang, P. Gao, Miscellaneous and perspicacious: hybrid halide perovskite materials based photodetectors and sensors, Adv. Opt. Mater. 8 (2020) 1–23.
[16] L. Su, Y. Zhu, X. Xu, H. Chen, Z. Tang, X. Fang, Back-to-back symmetric Schottky-type UVA photodetector based on ternary alloy BeZnO, J. Mater. Chem. C 6 (2018) 7776–7782.
[17] L. Su, Y. Zuo, J. Xie, Scalable manufacture of vertical p-GaN/n-SnO_2 heterostructure for self-powered ultraviolet photodetector, solar cell and dual-color light emitting diode, InfoMat 3 (2021) 598–610.
[18] P. Prajapat, D.K. Singh, G. Gupta, Growth of III-nitrides by molecular beam epitaxy: unconventional substrates for conventional semiconductors, Mater. Sci. Eng. B 295 (2023) 116574.
[19] P. Prajapat, P. Vashishtha, P. Goswami, G. Gupta, Fabrication of Sb_2Se_3-based high-performance self-powered visible-NIR broadband photodetector, Mater. Sci. Semicond. 169 (2023) 107873.
[20] P. Vashishtha, P. Goswami, P. Prajapat, A.K. Gangwar, P. Singh, G. Gupta, Highly efficient, self-powered, and air-stable broadband photodetector based on SnSe thin film, Mater. Sci. Eng. B 297 (2023) 116808.
[21] P. Prajapat, P. Vashishtha, P. Goswami, G. Gupta, Fabrication of Sb_2S_3/Sb_2Se_3 heterostructure for potential resistive switching applications, Nano Express 5 (2024) 015005.
[22] M. Long, P. Wang, H. Fang, W. Hu, Progress, challenges, and opportunities for 2D material based photodetectors, Adv. Funct. Mater. 29 (2019) 1–28.
[23] N. Huo, G. Konstantatos, Recent progress and future prospects of 2D-based photodetectors, Adv. Mater. 30 (2018) 1–27.
[24] S.K. Chamoli, G. Verma, S.C. Singh, C. Guo, Phase change material based hot electron photodetection, Nanoscale 13 (2021) 1311–1317.
[25] F. Aousgi, M. Kanzari, Structural and optical properties of amorphous Sb_2S_3 thin films deposited by vacuum thermal evaporation method, Curr. Appl. Phys. 13 (2013) 262–266.
[26] K. Ye, B. Wang, A. Nie, K. Zhai, F. Wen, C. Mu, Z. Zhao, J. Xiang, Y. Tian, Z. Liu, Broadband photodetector of high quality Sb_2S_3 nanowire grown by chemical vapor deposition, J. Mater. Sci. Technol. 75 (2021) 14–20.
[27] J. Chao, S. Xing, J. Zhang, C. Qin, D. Duan, X. Wu, Q. Shen, Synthesis of Sb_2S_3 nanowall arrays for high performance visible light photodetectors, Mater. Res. Bull. 57 (2014) 300–305.
[28] X. Lin, H. Deng, Y. Jia, Z. Wu, Y. Xia, X. Wang, S. Chen, Y. Cheng, Q. Zheng, Y. Lai, S. Cheng, S. Cheng, Self-powered Sb_2S_3 thin-film photodetectors with high detectivity for weak light signal detection, ACS Appl. Mater. Interfaces 14 (2022) 12385–12394.
[29] K. Zhao, J. Yang, M. Zhong, Q. Gao, Y. Wang, X. Wang, W. Shen, C. Hu, K. Wang, G. Shen, M. Li, J. Wang, W. Hu, Z. Wei, Direct polarimetric image sensor and wide spectral response based on quasi-1D Sb_2S_3 nanowire, Adv. Funct. Mater. 31 (2021) 1–11.
[30] J. Zhang, K. Hou, T. Huang, J. Yang, B. Xu, S. Li, B. Yang, W. Ma, Single-crystal Sb_2S_3 tube prepared by chemical vapor deposition for a 1 cm photodetector application, ACS Appl. Electron. Mater. 4 (5) (2022) 2455–2462.
[31] A. Bera, A. Das Mahapatra, S. Mondal, D. Basak, Sb_2S_3/Spiro-OMeTAD inorganic-organic hybrid p-n junction diode for high performance self-powered photodetector, ACS Appl. Mater. Interfaces 8 (2016) 34506–34512.
[32] J. Wu, N. Mao, L. Xie, H. Xu, J. Zhang, Identifying the crystalline orientation of black phosphorus using angle-resolved polarized Raman spectroscopy, Angew. Chem. - Int. Ed. 54 (2015) 2366–2369.
[33] I. Efthimiopoulos, C. Buchan, Y. Wang, Structural properties of Sb_2S_3 under pressure: evidence of an electronic topological transition, Sci. Rep. 6 (1–9) (2016) 24246.
[34] Y. Liu, K.T. Eddie Chua, T.C. Sum, C.K. Gan, First-principles study of the lattice dynamics of Sb_2S_3, Phys. Chem. Chem. Phys. 16 (2014) 345–350.
[35] X. Liu, Z. Feng, Y. Sun, M. Su, Y. Liu, H. Liu, J. Wen, H. Liu, Nanostructured CdS buffer layer fabricated with a simple spin-coating method for Sb_2S_3 solar cells, Phys. Status Solidi Appl. Mater. Sci. 218 (2021) 2100337, 1–7.
[36] P. Vashishtha, P. Prajapat, L. Goswami, A. Yadav, A. Pandey, G. Gupta, Stress-relaxed AlN-buffer-oriented GaN-nano-obelisks-based high-performance UV photodetector, Electron. Mater. 3 (2022) 357–367.
[37] J. Escorcia-García, M. Domínguez-Díaz, A. Hernández-Granados, H. Martínez, Antimony sulfide thin films obtained by chemical bath deposition using tartaric acid as complexing agent, MRS Adv. 3 (56) (2018) 1–7.
[38] H. Deng, S. Yuan, X. Yang, F. Cai, C. Hu, K. Qiao, J. Zhang, J. Tang, H. Song, Z. He, Efficient and stable TiO_2/Sb_2S_3 planar solar cells from absorber crystallization and Se-atmosphere annealing, Mater. Today Energy 3 (2017) 15–23.
[39] V. Lojpur, J. Krstić, Z. Kačarević-Popović, M. Mitrić, Z. Rakočević, I.L.J. Validžić, Efficient and novel Sb_2S_3 based solar cells with chitosan/poly(ethylene glycol)/electrolyte blend, Int. J. Energy Res. 42 (2018) 843–852.
[40] C.J. Diliegros-Godines, J. Santos Cruz, N.R. Mathews, M. Pal, Effect of Ag doping on structural, optical and electrical properties of antimony sulfide thin films, J. Mater. Sci. 53 (2018) 11562–11573.
[41] P. Vashishtha, P. Prajapat, A. Sharma, P. Goswami, S. Walia, G. Gupta, Self-driven and thermally resilient highly responsive nano-fenced MoS_2 based photodetector for near-infrared optical signal, Mater. Res. Bull. 164 (2023) 112260.
[42] P. Vashishtha, P. Prajapat, A. Sharma, P. Singh, S. Walia, G. Gupta, Self-Driven UVC–NIR broadband photodetector with high-temperature reliability based on a Coco palm-like MoS_2/GaN heterostructure, ACS Appl. Electron. Mater. 5 (3) (2023) 1891–1902.
[43] L. Luo, Y. Huang, K. Cheng, A. Alhassan, M. Alqahtani, L. Tang, Z. Wang, J. Wu, MXene-GaN van der Waals metal-semiconductor junctions for high performance multiple quantum well photodetectors, Light Sci. Appl. 10 (2021) 1–11.
[44] J. Zhu, L. Yu, Z. Wang, X. Wang, Y. Ou, M. Cai, Z. Wu, R. Tang, Y. Xia, Y. Wang, H. Deng, High-performance and stable Sb_2S_3 thin-film photodetectors for potential application in visible light communication, ACS Appl. Mater. Interfaces 15 (2023) 28175–28183.
[45] R. Li, H. Zeng, S. Bai, Z. Jia, Y. Xu, F. Yao, Y. Liu, H. Huang, Y. Li, R.J. Xie, Q. Lin, Modulating the quantum efficiency of Sb_2S_3-based photodiodes based on conventional and inverted structures, Laser Photonics Rev. 17 (2023) 2201006.
[46 K. Zhang, T. Luo, H. Chen, Z. Lou, G. Shen, Au-nanoparticles-decorated Sb_2S_3 nanowire-based flexible ultraviolet/visible photodetectors, J. Mater. Chem. C 5 (2017) 3330–3335.
[47] E. Aslan, M. Zarbali, Preparation of high-performance Sb_2S_3 based visible-light photodetector with excellent reversibility, Opt. Mater. 133 (2022) 113028.
[48] X. Lin, H. Deng, Y. Jia, Z. Wu, Y. Xia, X. Wang, S. Chen, Y. Cheng, Q. Zheng, Y. Lai, S. Cheng, Self-powered Sb_2S_3 thin-film photodetectors with high detectivity for weak light signal detection, ACS Appl. Mater. Interfaces 14 (10) (2022) 12385–12394.
[49] Y. Xiao, C. Li, X. Tan, L. Zhang, Y. Zhong, H. Zhu, Full-inorganic thin film solar cell and photodetector based on "graphene-on-antimony sulfide" heterostructure, Sol. RRL 1 (12) (2017) 1700135.



Mr. Pukhraj Prajapat is a Research Scholar & pursuing his Ph. D. from the Academy of Scientific & Innovative Research (AcSIR). He obtained his Master's degree in Physics from Jai Narayan Vyas University, Jodhpur, India. His current research activity is focused on the fabrication of metal di-chalcogenides (2D) based on optical and neuromorphic sensors.



Mr. Pargam Vashishtha obtained his M.Sc and M.Phil (Gold -Medalist) from CCS University, Meerut, India. He is pursuing a Joint (Cotutelle) PhD from the AcSIR, India and RMIT, Australia. He has authorised more than 85 SCI research articles the till now in the duration of his PhD.



Dr Deependra Kumar Singh is a DST-INSPIRE Faculty at CSIR-National Physical Laboratory in New Delhi. He holds a B. Tech. from IIT Roorkee, a Ph.D. from IISc Bangalore, and completed his postdoctoral work at the University of Dayton, USA. His work focus on exploring & understanding the basic electronic and optoelectronic properties of 2D/3D heterostructure-based devices for resistive sensing. He is also involved in neuromorphic computing using 2D materials, metal oxides, & III-nitrides.



Rajat Kumar Mukherjee is working as a Technical Officer at Optical Radiation Standards, at CSIR-National Physical Laboratory in New Delhi. He is pursuing his Ph.D. from the Academy of Scientific & Innovative Research (AcSIR).



Dr Jai S Tawale is working as Technical Officer at In-house BND Group, CSIR-National Physical Laboratory. New Delhi. He did his Ph.D (Physics) from NIT, Kurukshetra, India. He has published more than 40 research articles.



Dr. Lalit Goswami is working as Technical Officer at Sensor Devices & Metrology Group, CSIR-National Physical Laboratory. New Delhi He did his Ph.D. (Engineering Sciences) from Department of Electronics & Communication Engineering, Delhi Technological University, New Delhi. He has published more than 30 research artic;le and He is a fellow of Institute of Electronics & Telecommunication Engineers (FIETE).



Prof. Dr. Govind Gupta is Chief Scientist & Professor, Academy of Scientific & Innovative Research (AcSIR), India and Head, Sensor Devices & Metrology Group, CSIR-National Physical Laboratory (National Measurement Institute of India), New Delhi, India. His core area of expertize is the growth of III-nitrides, metal oxide & 2D materials, and the fabrication of smart optical & gas sensors. He has published $>$ 340 research articles in SCI journals. He is an elected fellow of the Royal Society of Chemistry (FRSC), Fellow, Institute of Physics (FInstP), UK, Fellow, Institute of Electronics & Telecommunication Engineers (FIETE), Fellow of Electron Microscopy Society of India, (FEMSI), Senior member of IEEE, Associate Academician of the Asia Pacific Academy of Materials, and received numerous awards & fellowships, including the MRSI-Distinguished Lectureship Award, MRSI medal, NASI-Young Scientist Medal, BOYSCAST fellowship, etc.

Gravitational Lensing and Clues of r_d and H_0 Tension: Viscous Modified Chaplygin Gas and Variable Modified Chaplygin Gas in Loop Quantum Cosmology

Rownak Kundu,[1,2,*] Ujjal Debnath,[2,†] Himanshu Chaudhary,[3,4,5,‡] and G. Mustafa[6,§]

[1]*Department of Mathematics, Nalbari College, Nalbari-781335, Assam, India.*
[2]*Department of Mathematics, Indian Institute of Engineering Science and Technology, Shibpur, Howrah-711 103, India.*
[3]*Department of Applied Mathematics, Delhi Technological University, Delhi-110042, India*
[4]*Pacif Institute of Cosmology and Selfology (PICS), Sagara, Sambalpur 768224, Odisha, India*
[5]*Department of Mathematics, Shyamlal College, University of Delhi, Delhi-110032, India.*
[6]*Department of Physics, Zhejiang Normal University, Jinhua 321004, People's Republic of China,*

arXiv:2402.07515v1 [gr-qc] 12 Feb 2024

This paper investigates the accelerated cosmic expansion in the late Universe by examining two dark energy models, namely viscous modified Chaplygin gas (VsMCG) and variable modified Chaplygin gas (VMCG), within the framework of loop quantum cosmology alongside the ΛCDM model. Our main objective is to constrain fundamental cosmic parameters in these dark energy models with ΛCDM model using 30 of the latest $H(z)$ measurements obtained from the cosmic chronometers method (CC), including Type Ia Supernovae, Gamma-Ray Bursts (GRB), Quasars, and 24 uncorrelated baryon acoustic oscillations (BAO) measurements from recent galaxy surveys across a redshift range from $0.106 < z < 2.33$. Additionally, we include the most recent Hubble constant measurement from Riess in 2022 to further enhance our constraints on fundamental cosmic parameters. In the ΛCDM, VsMCG, and VMCG frameworks, Our investigation yields best-fit parameters for the Hubble parameter (H_0) and sound horizon (r_d). The outcomes underscore a significant disparity between (H_0) and r_d) values using late-time observational measurements, reflecting the widely recognized H_0 and r_d tensions. Further, we study the gravitational lensing optical depth of the two dark energy models by plotting $\tau(z_s)/F$ vs z_s graphs. The probability of finding gravitational lenses (optical depth) in both our models increases with source redshift z_s. The change in optical depth behaviour for different sets of parameter value constraints through data sets is also graphically analysed. Joint analysis of VsMCG and VMCG with standard cosmological model ΛCDM is also carried out to study the optical depth behaviour. While the three models diverge in the early Universe, but for low redshift, they are found to be indistinguishable from each other. Finally, by employing the Akaike information criteria approach for model comparison, our analysis indicates that none of the two dark energy models can be dismissed based on the latest observational measurements.

I. INTRODUCTION

Our universe is undergoing accelerated expansion, a phenomenon supported by multiple independent studies [1–4] conducted over the past two decades. While the actual cause of this accelerated phenomenon is still debatable, many attribute it to huge negative pressure build-up in our present Universe. Many assume that some kind of 'mysterious' energy must be at play here and have named it 'dark energy' (DE, hereafter). Dark energy, or DE, almost occupies 70% of our Universe, so its study has been of immense importance. Over the years, researchers have proposed numerous such DE candidates to explain our accelerating Universe, and among them, the cosmological constant Λ happens to be the simplest. The other candidates includes Chaplygin gas [5, 6], quintessence [7, 8], phantom energy [9], holographic dark energy [10, 11], among the numerous others in literature [12]. Despite the diverse range of proposed dark energy candidates, the key to unravelling the mysteries of dark energy lies not only in theoretical constructs but also in empirical observations. Observational cosmology has thus become instrumental in discerning the nature of dark energy. Among the observational probes employed by cosmologists, the study of the large-scale structure of the Universe stands out prominently. Observations of the large-scale structure, which include the spatial distribution of galaxies, galaxy clusters, and cosmic voids, offer invaluable insights into the underlying dynamics of cosmic evolution [13]. By scrutinizing the cosmic web on vast scales, cosmologists can glean crucial information about the composition of the Universe, its expansion history, and the influence of dark energy on cosmic structures [14]. However, as observational techniques have advanced, they have revealed a perplexing tension between early-time and late-time cosmological observations.

* rownakkundu@gmail.com
† ujjaldebnath@gmail.com
‡ himanshuch1729@gmail.com
§ gmustafa3828@gmail.com

This tension manifests in discrepancies between measurements obtained from the early Universe, such as those derived from the cosmic microwave background (CMB) radiation [15] and those obtained from more recent observations of the local Universe, such as those from Type Ia supernovae [16] and Baryon Acoustic Oscillations (BAO). BAO plays a pivotal role, offering a unique and powerful tool for understanding the large-scale structure of the Universe. These oscillations are imprints left on the distribution of matter in the early Universe, originating from the coupled interaction between photons and baryons (protons and neutrons) before the era of recombination [17]. The significance of BAO lies in their ability to serve as standard rulers on cosmological scales. As the Universe expands, these primordial density fluctuations lead to characteristic patterns in the distribution of galaxies and cosmic microwave background radiation. The formation of BAO can be traced back to the sound waves that propagated through the primordial plasma of the early Universe. As the Universe expanded, the sound waves left behind a distinctive pattern of overdense and underdense regions, creating a preferred scale known as the BAO scale or sound horizon [18]. This scale acts as a standard ruler, allowing cosmologists to measure the geometry and expansion rate of the Universe over cosmic time. The BAO scale, or sound horizon, serves as a crucial cosmological standard because its size is determined by fundamental cosmological parameters, such as the density of baryons and dark matter, as well as the speed of sound in the primordial plasma. Observations of BAO in the large-scale distribution of galaxies provide a powerful constraint on these parameters and offer insights into the nature of dark energy [19]. In the field of cosmology, the study of BAO and the sound horizon has become an essential tool for precision cosmological measurements. Large galaxy surveys, such as the Sloan Digital Sky Survey (SDSS), have played a key role in mapping the distribution of galaxies and detecting the BAO signal [20]. By analyzing the BAO scale in the clustering of galaxies, researchers can infer the expansion history of the Universe and shed light on the mysterious components of dark matter and dark energy. Thus, Baryon Acoustic Oscillations stand as a cornerstone in our quest to unravel the fundamental properties and evolution of the cosmos.

While BAO provide valuable insights about the large-scale structure of the Universe and the behavior of dark energy, the accelerating expansion of the cosmos remains a profound puzzle that has spurred researchers to explore alternative explanations. Many assume that Einstein's gravitational theory might be incomplete. So, researchers modified the underlying gravity theory (and hence, the name modified theory of gravity [21, 22]) to incorporate the accelerating phase of our Universe. While these theories effectively explain the formation of structures and gravitational interactions on a large scale, quantum gravity becomes necessary to understand behaviors at smaller scales. Further, the backward evolution of our Universe in time results in the collapse of our Universe into a single point with diverging energy density. Such instances lead to the failure of our classical gravity theory, rendering it ineffective in describing the unfolding events. Quantum gravity, with its distinct dynamics on smaller scales, is expected to solve this dilemma. One such theory based upon quantum gravity is loop quantum cosmology (LQC, hereafter) [23, 24]. In recent years, several DE models have been studied in its framework. Jamil et al. [25] investigated the combination of modified Chaplygin gas with dark matter within the LQC framework, resolving the cosmic coincidence problem. The authors in ref. [26], discovered that loop quantum effects could prevent future singularities in the FRW cosmology. Given its importance, in our present study as well, we considered two models of Chaplygin gas, namely, viscous modified Chaplygin gas (VsMCG) and variable modified Chaplygin gas (VMCG) in LQC framework, to analyze their optical depth behaviour by constraining the model parameters to their best-fit values. Further, the significance of the study of gravitational lensing also has long been recognised in cosmology, and the phenomenal work of Refsdal et al. [27] in this context is worth mentioning. Over recent years, gravitational lensing emerged as an important tool for studying dark matter, dark energy, and black holes, among many others. Gravitational lensing optical depth represents the likelihood of the formation of multiple images owing to the influence of gravitational lenses in our Universe. It was first used by ref. [28, 29] and consequently by many others [30–32], including Kundu et al. [33] to study and understand dark energy models. Our present paper uses this lensing phenomenon to qualitatively analyse the two DE models against their constraint parameter values and record the changes graphically. This methodology of studying the gravitational lensing optical depth of DE models can give us an estimation in understanding the large-scale structure of our Universe, thereby helping us better understand our models.

Finally, our paper is organized as follows: Section II delves into the background equations of the two dark energy models within the framework of loop quantum

cosmology. Section III outlines the methodology employed to constrain the parameters of the dark energy models using various datasets. Section IV is dedicated to the study of the lensing phenomenon, including the derivation of the equation for optical depth to be used in our study. In Section V, we present the outcomes of our study, and we conclude by discussing our findings in Section VI.

II. BACKGROUND EQUATIONS: LOOP QUANTUM COSMOLOGY

In loop quantum cosmology, for a isotropic and homogeneous Universe the background equations are defined by:

$$H^2 = \frac{1}{3}\left(1 - \frac{\rho}{\rho_1}\right)\rho, \tag{1}$$

$$\dot{H} = -\frac{1}{2}\left(1 - \frac{2\rho}{\rho_1}\right)(\rho + p) \tag{2}$$

where $\rho_1 = \frac{\sqrt{3}}{16\pi^2\gamma^3 G^2\hbar}$ is the critical loop quantum density and γ the dimensionless "Barbero-Immirzi parameter". The conservation equation in LQC is given by:

$$\dot{\rho} + 3H(\rho + p) = 0 \tag{3}$$

Assuming the matter content of our Universe to be a combination of DM and DE, the quantities ρ and p defined in Eqs (1) and (2) will modify to $\rho = \rho_m + \rho_d$ and $p = p_m + p_d$. With the consideration that both DM and DE are separately conserved, the equation (3) now seperates into two distinct equations as follows:

$$\dot{\rho}_m + 3H(\rho_m + p_m) = 0, \tag{4}$$

$$\dot{\rho}_d + 3H(\rho_d + p_d) = 0, \tag{5}$$

Since, pressure in dark matter is very negligible (i.e. $p_m = 0$), equation (4) gives $\rho_m = \rho_{m0}(1+z)^3$. Here, ρ_{m0} represent the density value at the present epoch.

A. Viscous modified Chaplygin gas (VsMCG)

The expression for pressure in the case of viscous modified Chaplygin gas (VsMCG) is given by [34]:

$$p_d = A\rho_d - \frac{B}{\rho_d^{\alpha}} - 3\zeta_0\sqrt{\rho_d}H \tag{6}$$

where A, B and α are constants. Substituting p_d from Eq (6) in Eq (5), we get:

$$\rho_d = \left(\frac{B}{1 + A - \sqrt{3}\zeta_0} + \frac{C_1}{a^{3(1+\alpha)(1+A-\sqrt{3}\zeta_0)}}\right)^{\frac{1}{1+\alpha}} \tag{7}$$

where C_1 is an integrating constant. The above expression can be further re-written as:

$$\rho_d = \rho_{d0}\{A_s + (1 - A_s)(1+z)^{3(1+\alpha)(1+A-\sqrt{3}\zeta_0)}\}^{\frac{1}{1+\alpha}} \tag{8}$$

where ρ_{d0} being the DE density value at the present epoch, $A_s = \frac{B}{(1+A-\sqrt{3}\zeta_0)C_1+B}$ satisfying the conditions $0 < A_s < 1$ and $1 + A - \sqrt{3}\zeta_0 > 0$, and $\rho_{d0}^{1+\alpha} = \frac{(1+A-\sqrt{3}\zeta_0)C_1+B}{1+A-\sqrt{3}\zeta_0}$. Thus, we can write the Hubble parameter from equation (1) as follows:

$$H^2(z) = H_0^2\left[\Omega_{m0}(1+z)^3 + \Omega_{d0}\left\{A_s + (1-A_s)(1+z)^{3(1+\alpha)(1+A-\sqrt{3}\zeta_0)}\right\}^{\frac{1}{1+\alpha}}\right]\times \left[1 - \frac{3H_0^2}{\rho_1}\left\{\Omega_{m0}(1+z)^3 + \Omega_{d0}\left\{A_s + (1-A_s)(1+z)^{3(1+\alpha)(1+A-\sqrt{3}\zeta_0)}\right\}^{\frac{1}{1+\alpha}}\right\}\right] \tag{9}$$

where $\Omega_{m0} = \frac{\rho_{m0}}{3H_0^2}$ and $\Omega_{d0} = \frac{\rho_{d0}}{3H_0^2}$ are the dimensionless parameters all defined at the present epoch. Further, substituting $z = 0$ in equation (9) and equating the RHS to '1' we get:

$$\left[\Omega_{m0} + \Omega_{d0}\right] \times \left[1 - \frac{3H_0^2}{\rho_1}\left\{\Omega_{m0} + \Omega_{d0}\right\}\right] = 1 \tag{10}$$

which simplifies to

$$\Omega_{d0} = \frac{1}{6H_0^2}\left(\rho_1 + \sqrt{\rho_1^2 - 12\rho_1 H_0^2}\right) - \Omega_{m0} \tag{11}$$

B. Variable modified Chaplygin gas (VMCG)

The expression for pressure in the case of variable modified Chaplygin gas (VMCG) is given by [35]:

$$p_d = C\rho_d - \frac{D(a)}{\rho_d^\beta} \tag{12}$$

where $0 \leq \beta \leq 1$ and $C > 0$. Assuming, $D(a) = D_0 a^{-n}$ where $D_0 > 0$ and $n > 0$, we get from Eq (5):

$$\rho_d = \left[\frac{3(1+\beta)D_0}{\{3(1+\beta)(1+C)-n\}a^n} + \frac{K}{a^{3(1+C)(1+\beta)}}\right]^{\frac{1}{1+\beta}} \tag{13}$$

where $K(>0)$ is an arbitrary integrating constant and $3(1+\beta)(1+C) - n > 0$. Also, here n must be positive, or else for $a \to \infty$ we would have $\rho_d \to \infty$ thereby contradicting our case for expanding Universe. The expression for ρ_d above can be further simplified to:

$$\rho_d = \rho_{d0}\left[\frac{C_s}{a^n} + \frac{1-C_s}{a^{3(1+C)(1+\beta)}}\right]^{\frac{1}{1+\beta}} \tag{14}$$

where ρ_{d0} denotes the DE density value at the present epoch, $C_s = 1 - \frac{K}{\rho_{d0}^{1+\beta}}$ and $\rho_{d0}^{1+\beta} = K + \frac{3(1+\beta)D_0}{3(1+\beta)(1+C)-n}$. Thus, we can write the Hubble parameter from equation (1) as follows:

$$H^2(z) = H_0^2\left[\Omega_{m0}(1+z)^3 + \Omega_{d0}\left\{C_s(1+z)^n + (1-C_s)(1+z)^{3(1+C)(1+\beta)}\right\}^{\frac{1}{1+\beta}}\right] \times \left[1 - \frac{3H_0^2}{\rho_1}\left\{\Omega_{m0}(1+z)^3 + \Omega_{d0}\left\{C_s(1+z)^n + (1-C_s)(1+z)^{3(1+C)(1+\beta)}\right\}^{\frac{1}{1+\beta}}\right\}\right] \tag{15}$$

where Ω_{d0} is given in (11).

III. METHODOLOGY

In our study, we carefully picked a specific set of recent Baryon Acoustic Oscillation (BAO) measurements from various galaxy surveys, with the primary contributions coming from observations made by the Sloan Digital Sky Survey (SDSS) [36–41]. We also included valuable data from the Dark Energy Survey (DES) [42], the Dark Energy Camera Legacy Survey (DECaLS) [43], and 6dFGS BAO [44] to enhance the diversity of our dataset. Recognizing the potential for correlations among selected data points, we took steps to address this issue. While we intentionally crafted a subset to minimize highly correlated points, acknowledging and dealing with possible correlations within our chosen data points was deemed crucial. For a robust estimation of systematic errors, we used mocks generated from N-body simulations to accurately determine covariance matrices. Given the diverse nature of measurements from various observational surveys, obtaining precise covariance matrices between them posed a significant challenge. To address this, we adopted the covariance analysis approach outlined in [45], specifically using $C_{ii} = \sigma_i^2$. To simulate correlations within our chosen subsample, we introduced non-diagonal elements into the covariance matrix while maintaining symmetry. Incorporating non-negative correlations involved randomly selecting up to 12 pairs of data points and assigning non-diagonal elements with magnitudes of $C_{ij} = 0.5\sigma_i\sigma_j$. Here, σ_i and σ_j denote the 1σ errors of data points i and j. This approach allowed us to effectively represent correlations within 55% of our chosen BAO dataset. The variation of some cosmological parameters of ΛCDM according to the number of correlated pairs is shown in Table I. Fig. 1 displays the posterior distributions for the ΛCDM Model with and without the incorporation of a test random covariance matrix comprising fourteen components. The influence of this covariance matrix, ranging from null to fourteen

Parameters	n correlated pairs	BAO	BAO+R22
Ω_{m0}	n = 0	0.270 ± 0.015	0.266 ± 0.015
$\Omega_{\Lambda 0}$		0.725 ± 0.011	0.728 ± 0.0125
Ω_{m0}	n = 7	0.264 ± 0.016	0.261 ± 0.015
$\Omega_{\Lambda 0}$		0.730 ± 0.013	0.733 ± 0.011
Ω_{m0}	n = 14	0.266 ± 0.015	0.264 ± 0.016
$\Omega_{\Lambda 0}$		0.728 ± 0.011	0.730 ± 0.012

TABLE I: The table shows slight variations in different cosmological parameters when comparing scenarios with uncorrelated pairs ($n = 0$) to those with $n = 7$ and $n = 14$ randomly introduced correlated pairs.

components, appears negligible, closely resembling the uncorrelated dataset. Moving on to determining the best-fit values for our cosmological model parameters, we expanded the BAO dataset by incorporating thirty uncorrelated Hubble parameter measurements obtained through the cosmic chronometers (CC) method discussed in [46–49]. Additionally, we included the latest Pantheon sample data on Type Ia Supernovae [50]. To broaden our observational scope further, we introduced 24 binned quasar distance modulus data from [51], a set of 162 Gamma-Ray Bursts (GRBs) as outlined in [52], and the recent Hubble constant measurement (R22) [53] as an additional prior. In our analysis, we used a nested sampling approach implemented in the open-source Polychord package [54], complemented by the GetDist package [55] to present our results in a clear and informative manner.

IV. GRAVITATIONAL LENSING

The differential probability that a distribution of galaxies per unit redshift will be multiply imaged by a source with redshift z_s is [33]:

$$\frac{d\tau}{dz_l} = n(1+z_l)^3 S \frac{cdt}{dz_l} \tag{16}$$

where n denotes the co-moving number density of lenses, S is the lens cross-section and cdt/dz_l the proper distance interval. Further, following numerous work [29, 56] done previously, we assume an singular isothermal sphere (SIS) profile for our lensing model. Thus, the expression for lensing cross-section in an SIS profile is given by:

$$S = 16\pi^3 \left(\frac{\sigma}{c}\right)^4 \left(\frac{D_l D_{ls}}{D_s}\right)^2 \tag{17}$$

where σ is the velocity dispersion, D_s, D_l and D_{ls} represents the angular diameter distances between observer-source, observer-lens and lens-source respectively, given by:

$$D_{z_1 z_2} = \frac{1}{\sqrt{H_0}(1+z_2)} \int_{z_1}^{z_2} \frac{1}{\sqrt{H(z)}} dz \qquad (k=0) \tag{18}$$

and finally the proper distance interval reads:

$$\frac{cdt}{dz_l} = \frac{c}{(1+z_l)H(z_l)} \tag{19}$$

Now, in order to calculate n present in Eq. (16), we use the Schechter function given by:

$$\frac{dn}{dL} = \frac{n_i^*}{L_i^*} \left(\frac{L}{L_i^*}\right)^{\gamma_i} e^{(-L/L_i^*)} \tag{20}$$

where i depends on the type of galaxy morphology (i.e. S, S_0 or E, n_i^* denotes the characteristic number density, and l_i^* the characteristic luminosity. Additionally, with the help of the power-law relation $L/L_i^* = (\sigma/\sigma_i^*)^{g_i}$, equation (20) further reduces to

$$\frac{dn}{d\sigma} = \frac{n_i^*}{\sigma_i^*} \left(\frac{\sigma}{\sigma_i^*}\right)^{g_i\gamma_i + g_i - 1} e^{(-\sigma/\sigma_i^*)^{g_i}} g_i \tag{21}$$

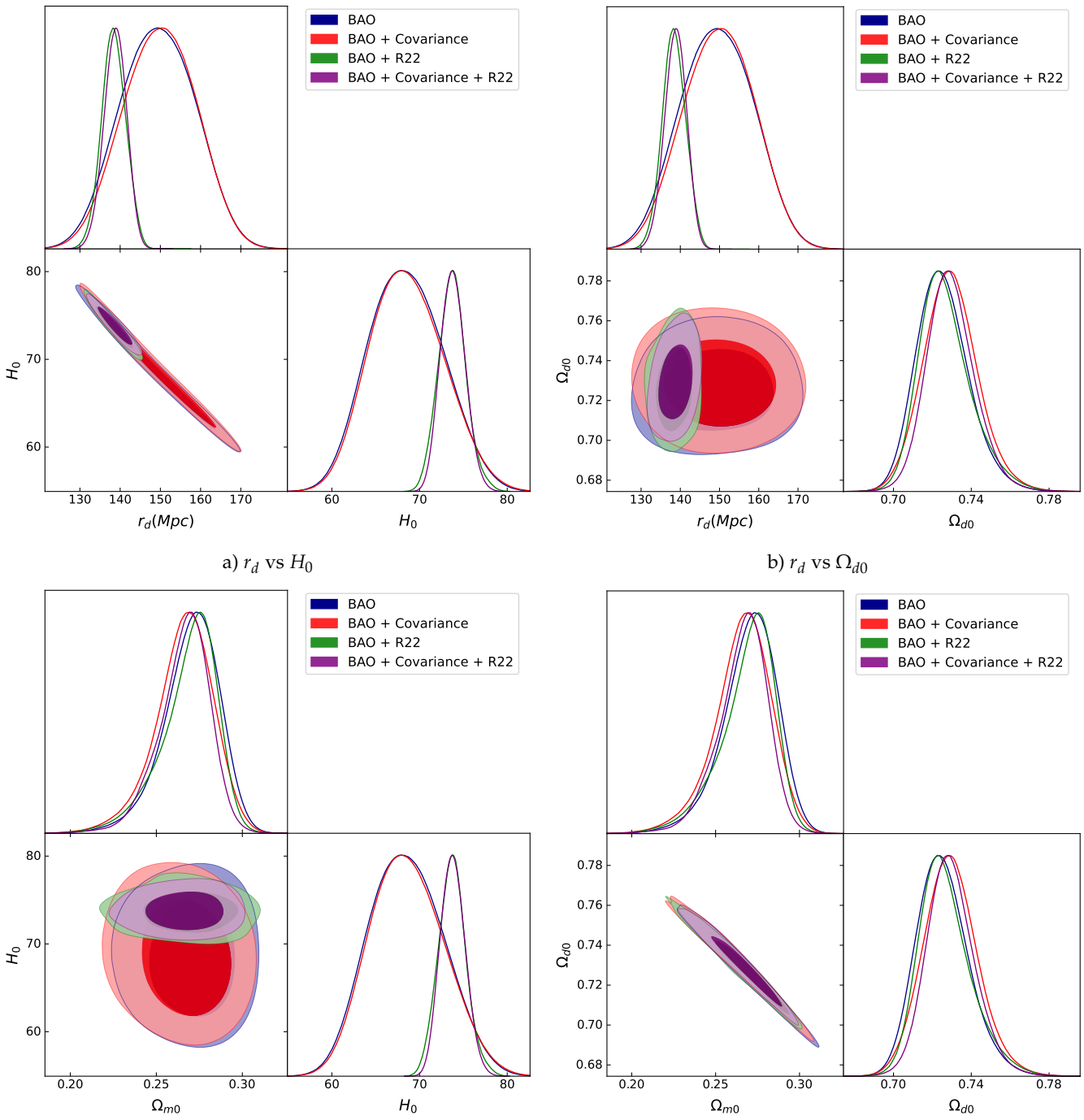
Thus, the differential optical depth can now be obtained by integrating eqn. (16) from 0 to z_s as follows:

$$\tau(z_s) = \left(\sum_{i=S,S_0,E} F_i\right) \int_0^{\chi_s} \left(\frac{\chi_s - \chi_l}{\chi_s}\right)^2 \chi_l^2 d\chi_l \tag{22}$$

where

$$\chi_i = \sqrt{H_0} \int_0^{z_i} \frac{dz}{\sqrt{H(z)}} \tag{23}$$

is the co-moving distances, F_i represents the ability of the ith class of galaxy morphology to produce multiple images, relying exclusively on the intrinsic and statistical properties of the galaxies. Using equation (21), we



a) r_d vs H_0

b) r_d vs Ω_{d0}

c) H_0 vs Ω_{d0}

d) Ω_{m0} vs Ω_{d0}

FIG. 1: Comparison of posterior distributions in ΛCDM with and without a randomly tested covariance matrix of fourteen components shows negligible impact. This renders the dataset nearly indistinguishable from an uncorrelated one. Covariance matrices beyond fourteen components exhibit minimal influence.

may calculate F_i in the form:

$$F_i = 16\pi^3\sigma^4 H_0^{-3} \int_0^\infty \frac{n_i^*}{\sigma_i^*} \left(\frac{\sigma}{\sigma_i^*}\right)^{g_i\alpha_i+g_i-1} e^{(-\sigma/\sigma_i^*)^{g_i}} g_i d\sigma$$
$$= 16\pi^3\sigma^4 H_0^{-3} n_i^* \Gamma(\alpha_i+1) \tag{24}$$

Considering the uncertainties extensively addressed in the literature [57–60], we will treat F to be a normalized factor, denoted as $F \equiv \sum_{i=S,S_0,E} F_i$. Therefore, from equations (22) and (24), we deduce a straightforward an-

MCMC Results						
Model	Parameters	Priors	BAO	BA0 + R22	CC + SC + BAO	CC + SC + BAO + R22
ΛCDM Model	H_0	[50,100]	$69.089209^{\pm 6.252595}_{\pm 4.396874}$	$73.906596^{\pm 2.788642}_{\pm 1.357697}$	$69.854848^{\pm 2.386935}_{\pm 1.259100}$	$71.616475^{\pm 1.936192}_{\pm 1.003940}$
	Ω_{m0}	[0.,1.]	$0.256796^{\pm 0.069423}_{\pm 0.025772}$	$0.254859^{\pm 0.068716}_{\pm 0.026912}$	$0.268654^{\pm 0.028134}_{\pm 0.012822}$	$0.264407^{\pm 0.030224}_{\pm 0.013179}$
	Ω_{d0}	[0.,1.]	$0.734631^{\pm 0.038211}_{\pm 0.020888}$	$0.736400^{\pm 0.034343}_{\pm 0.021510}$	$0.724585^{\pm 0.016999}_{\pm 0.009373}$	$0.728562^{\pm 0.018197}_{\pm 0.009532}$
	r_d (Mpc)	[100,200]	$149.504236^{\pm 15.212831}_{\pm 10.037166}$	$139.447508^{\pm 5.883881}_{\pm 2.914639}$	$146.543556^{\pm 5.101856}_{\pm 2.598566}$	$143.299915^{\pm 4.358875}_{\pm 2.218062}$
	r_d/r_{fid}	[0.9,1.1]	$1.008382^{\pm 0.101188}_{\pm 0.066344}$	$0.940599^{\pm 0.037308}_{\pm 0.020806}$	$0.990266^{\pm 0.035810}_{\pm 0.019826}$	$0.967284^{\pm 0.030444}_{\pm 0.015101}$
VsMCG Model	H_0	[50,100]	$65.168094^{\pm 6.617797}_{\pm 10.434466}$	$73.07201^{\pm 0.978415}_{\pm 1.889797}$	$67.447204^{\pm 2.514723}_{\pm 1.878032}$	$72.860428^{\pm 1.040048}_{\pm 2.328246}$
	Ω_{m0}	[0.,1.]	$0.365902^{\pm 0.069461}_{\pm 0.115990}$	$0.305203^{\pm 0.041952}_{\pm 0.071096}$	$0.346451^{\pm 0.069036}_{\pm 0.105164}$	$0.289643^{\pm 0.012301}_{\pm 0.025106}$
	Ω_{d0}	[0.,1.]	$0.809753^{\pm 0.133984}_{\pm 0.234411}$	$0.676112^{\pm 0.101291}_{\pm 0.190216}$	$0.779877^{\pm 0.158516}_{\pm 0.239042}$	$0.652845^{\pm 0.028308}_{\pm 0.059635}$
	A_s	[0.,1.]	$0.514312^{\pm 0.491146}_{\pm 0.279102}$	$0.504348^{\pm 0.457075}_{\pm 0.324263}$	$0.436312^{\pm 0.416373}_{\pm 0.319504}$	$0.447568^{\pm 0.430062}_{\pm 0.357146}$
	α	[0.,1.]	$0.683658^{\pm 0.445186}_{\pm 0.232932}$	$0.628384^{\pm 0.440275}_{\pm 0.262042}$	$0.665943^{\pm 0.144111}_{\pm 0.064040}$	$0.667478^{\pm 0.157306}_{\pm 0.074594}$
	A	[2.,4.]	$3.089772^{\pm 1.049832}_{\pm 0.649737}$	$2.965524^{\pm 0.891859}_{\pm 0.602096}$	$3.014177^{\pm 0.952994}_{\pm 0.699794}$	$3.093859^{\pm 1.043732}_{\pm 0.615597}$
	ζ_0	[1.5,2.5]	$1.962968^{\pm 0.442674}_{\pm 0.310779}$	$1.990460^{\pm 0.453422}_{\pm 0.349834}$	$2.101817^{\pm 0.555898}_{\pm 0.297352}$	$2.050346^{\pm 0.511089}_{\pm 0.357694}$
	ρ_1	[1310000,1390000]	$1351709.4^{\pm 38561.9}_{\pm 28642.8}$	$1349170.2^{\pm 36390.3}_{\pm 26320.8}$	$1349393.1^{\pm 37491.8}_{\pm 26568.3}$	$1353509.1^{+41653.0}_{-25630.2}$
	r_d (Mpc)	[100,200]	$149.571115^{\pm 16.426449}_{\pm 8.752594}$	$140.757158^{\pm 12.992500}_{\pm 8.454083}$	$147.559044^{\pm 5.811905}_{\pm 2.723020}$	$141.158148^{\pm 4.709648}_{\pm 2.299581}$
	r_d/r_{fid}	[0.9,1.1]	$1.009424^{\pm 0.104352}_{\pm 0.059881}$	$0.990016^{\pm 0.083577}_{\pm 0.056765}$	$0.995630^{\pm 0.039894}_{\pm 0.019954}$	$0.993363^{\pm 0.031652}_{\pm 0.016175}$
VMCG Model	H_0	[50,100]	$65.796665^{\pm 6.461974}_{\pm 10.070874}$	$73.799085^{\pm 1.325015}_{\pm 2.574862}$	$69.204717^{\pm 2.413623}_{\pm 1.056031}$	$73.025435^{\pm 1.390057}_{\pm 2.671485}$
	Ω_{m0}	[0.,1.]	$0.286515^{\pm 0.075623}_{\pm 0.172756}$	$0.197851^{\pm 0.052402}_{\pm 0.100870}$	$0.229029^{\pm 0.065811}_{\pm 0.145501}$	$0.181731^{\pm 0.060864}_{\pm 0.119488}$
	Ω_{d0}	[0.,1.]	$0.700150^{\pm 0.113069}_{\pm 0.186913}$	$0.803863^{\pm 0.129291}_{\pm 0.276353}$	$0.801456^{\pm 0.120813}_{\pm 0.212190}$	$0.717771^{\pm 0.067980}_{\pm 0.116377}$
	n	[0.5,1.5]	$0.990322^{\pm 0.007463}_{\pm 0.034757}$	$0.980549^{\pm 0.017220}_{\pm 0.048783}$	$0.976961^{\pm 0.018299}_{\pm 0.051484}$	$0.970191^{\pm 0.026968}_{\pm 0.065479}$
	β	[0.8,1.4]	$1.018741^{\pm 0.086082}_{\pm 0.115003}$	$1.020759^{\pm 0.084889}_{\pm 0.112896}$	$1.021356^{\pm 0.085708}_{\pm 0.115778}$	$1.025479^{\pm 0.097275}_{\pm 0.123108}$
	C	[0.,0.3]	$0.082870^{\pm 0.057201}_{\pm 0.079413}$	$0.082870^{\pm 0.057201}_{\pm 0.079413}$	$0.045065^{\pm 0.036463}_{\pm 0.044343}$	$0.038790^{\pm 0.033014}_{\pm 0.037825}$
	C_s	[0.,1.]	$0.599738^{\pm 0.361244}_{\pm 0.545885}$	$0.599738^{\pm 0.361244}_{\pm 0.545885}$	$0.376236^{\pm 0.293035}_{\pm 0.364827}$	$0.481279^{\pm 0.380702}_{\pm 0.472094}$
	ρ_1	[1310000,1390000]	$1351317.6^{\pm 27841.6}_{\pm 39991.2}$	$1349811.1^{\pm 25642.0}_{\pm 36678.6}$	$1354908.4^{\pm 26738.7}_{\pm 42104.6}$	$1350850.4^{\pm 27481.4}_{\pm 37505.8}$
	r_d (Mpc)	[100,200]	$148.683108^{\pm 14.573363}_{\pm 8.949488}$	$139.365379^{\pm 14.205048}_{\pm 9.762761}$	$147.906003^{\pm 4.444459}_{\pm 2.386071}$	$141.623321^{\pm 5.305695}_{\pm 2.638046}$
	r_d/r_{fid}	[0.9,1.1]	$1.002078^{\pm 0.097008}_{\pm 0.063968}$	$1.000626^{\pm 0.094807}_{\pm 0.067223}$	$0.995166^{\pm 0.030789}_{\pm 0.016201}$	$0.994314^{\pm 0.035566}_{\pm 0.019600}$

TABLE II: The table presents constraints on cosmological parameters for the ΛCDM, VsMCG, and VMCG models at a 95% confidence level.

alytical expression for the optical depth of a point source at redshift z_s in an FRW Universe with dark matter and dark energy components [33]:

$$\tau(z_s) = \frac{F}{30}\chi_s^3 \tag{25}$$

V. RESULTS

For the standard ΛCDM model, the posterior distribution of the key cosmological parameters is shown at the 68% and 95% confidence levels in Fig. 2. Detailed results from MCMC simulations are outlined in Table II. When the R22 prior is incorporated into the joint dataset, the optimal fitting value for H_0 is 71.50 ± 0.882, diverging [15] from but aligning closely with measurements from the SNIe sample in [53]. In the absence of R22 priors with the Joint dataset, the estimated $H_0 = 69.837 \pm 1.204$ corresponds more closely to the value in [15]. Figs 3 and 4 portray the 68% and 95% confidence levels for crucial cosmological parameters in VsMCG and VMCG models. Across these models, when the R22 prior is included in the Joint dataset, the optimal fitting value for H_0 diverges from [15] but

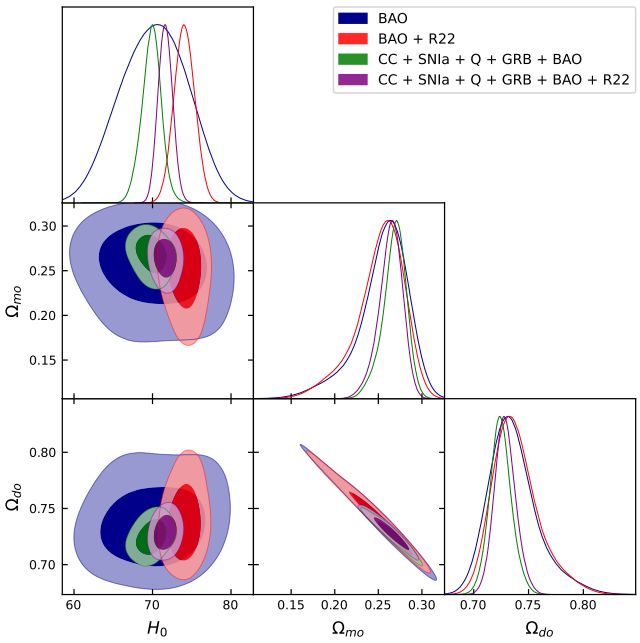


FIG. 2: The figure illustrates 1σ and 2σ regions using the standard ΛCDM model to demonstrate the posterior distribution of various observational data measurements.

aligns more closely with the SNIe sample in [53]. Conversely, without R22 priors and using the Joint dataset, the estimated $H_0 = 69.837 \pm 1.204$ aligns more closely with [15]. The determined values for the matter density, denoted as Ω_{m0}, and the dark energy density, denoted as Ω_{d0}, in the ΛCDM and VMCG Models, appear to be lower compared to the values documented in [15] ($\Omega_{m0} = 0.315 \pm 0.007$, $\Omega_{d0} = 0.685 \pm 0.007$). However, in the case of the VsMCG Model, it is higher than the value documented in [15]. However, this observation has been documented in alternative studies [61, 62]. In the context of the Baryon Acoustic Oscillations (BAO) scale, it is defined by the cosmic sound horizon imprinted in the cosmic microwave background during the drag epoch, denoted as z_d. This epoch marks the separation of baryons and photons. The BAO scale, represented by r_d, is determined by the integral of the ratio of the speed of sound (c_s) to the Hubble parameter (H_0) over the redshift range from z_d to infinity. The speed of sound, c_s, is given by $\sqrt{\frac{\delta p_\gamma}{\delta\rho_B+\delta\rho_\gamma}}$, where δp_γ is the pressure perturbation in photons, and $\delta\rho_B$ and $\delta\rho_\gamma$ are perturbations in the baryon and photon energy densities, respectively. This expression is further simplified to $\frac{1}{\sqrt{3(1+R)}}$, with R defined as the ratio of baryon density perturbation to photon density perturbation ($R \equiv \frac{\delta\rho_B}{\delta\rho_\gamma} = \frac{3\rho_B}{4\rho_\gamma}$). Observational data from [15] provides the redshift at the drag epoch as $z_d = 1059.94 \pm 0.30$. In the case of a ΛCDM model, measurements from [15] estimate the BAO scale r_d to be 147.09 ± 0.26 megaparsecs (Mpc). In our exploration of the ΛCDM model, Fig 5a illustrates the posterior distribution of the contour plane for $r_d - H_0$. The BAO datasets provide an estimated r_d of 145.807 ± 9.06 Mpc, aligning with the reported findings in [63]. However, when we exclusively incorporate R22 prior into the BAO dataset, the determined sound horizon at the drag epoch is 138.345 ± 2.45 Mpc. Examining the Joint dataset reveals an estimated BAO scale (r_d) of 145.811 ± 2.34 Mpc, closely aligning with the outcomes reported in [15]. Additionally, incorporating the R22 prior into the comprehensive dataset results in r_d of 142.591 ± 1.49 Mpc, indicating proximity to the findings in [64]. Now, shifting to the VsMCG Model, Fig 5b presents the posterior distribution for the $r_d - H_0$ contour plane. For the BAO datasets, the resulting r_d is 149.571 ± 8.752 Mpc, consistent with the Planck results as reported in [63]. However, incorporating R22 prior exclusively into the BAO dataset leads to a sound horizon at the drag epoch of 140.757 ± 8.454 Mpc. In the case of the Joint dataset, the determined r_d is 147.559 ± 2.72 Mpc, aligning closely with the Planck results. Furthermore, integrating the R22 prior into the full dataset results in $r_d = 141.158 \pm 2.299$ Mpc, showing proximity to the findings presented in [64]. In the context of the VMCG Model, Fig 5c illustrates the posterior distribution of the $r_d - H_0$ contour plane. When considering BAO datasets, the estimated BAO scale (r_d) is 148.683 ± 8.949 Mpc, a result consistent with the findings reported in [63]. However, introducing the R22 prior into the BAO dataset alone yields a sound horizon at the drag epoch of 139.365 ± 9.762 Mpc. In the case of the Joint dataset, the calculated BAO scale (r_d) is 147.906 ± 2.386 Mpc, which closely matches the results obtained from the Planck mission. Additionally, when we include the R22 prior in the complete dataset, the resulting r_d is 141.623 ± 2.638 Mpc, indicating a close agreement with the findings reported in [64]. The results from ΛCDM, VsMCG and VMCG models exhibit tension with the r_d value estimated by Planck. However, these results still demonstrate agreement with the findings presented in [64] reports that by employing Binning and Gaussian methods to combine measurements of 2D BAO and SNIa data, the values of the absolute BAO scale range from 141.45 Mpc to $r_d \leq 159.44$Mpc (Binning) and 143.35 Mpc to $r_d \leq 161.59$Mpc (Gaussian). These findings highlight a clear discrepancy between early and late-time observational measurements, analogous

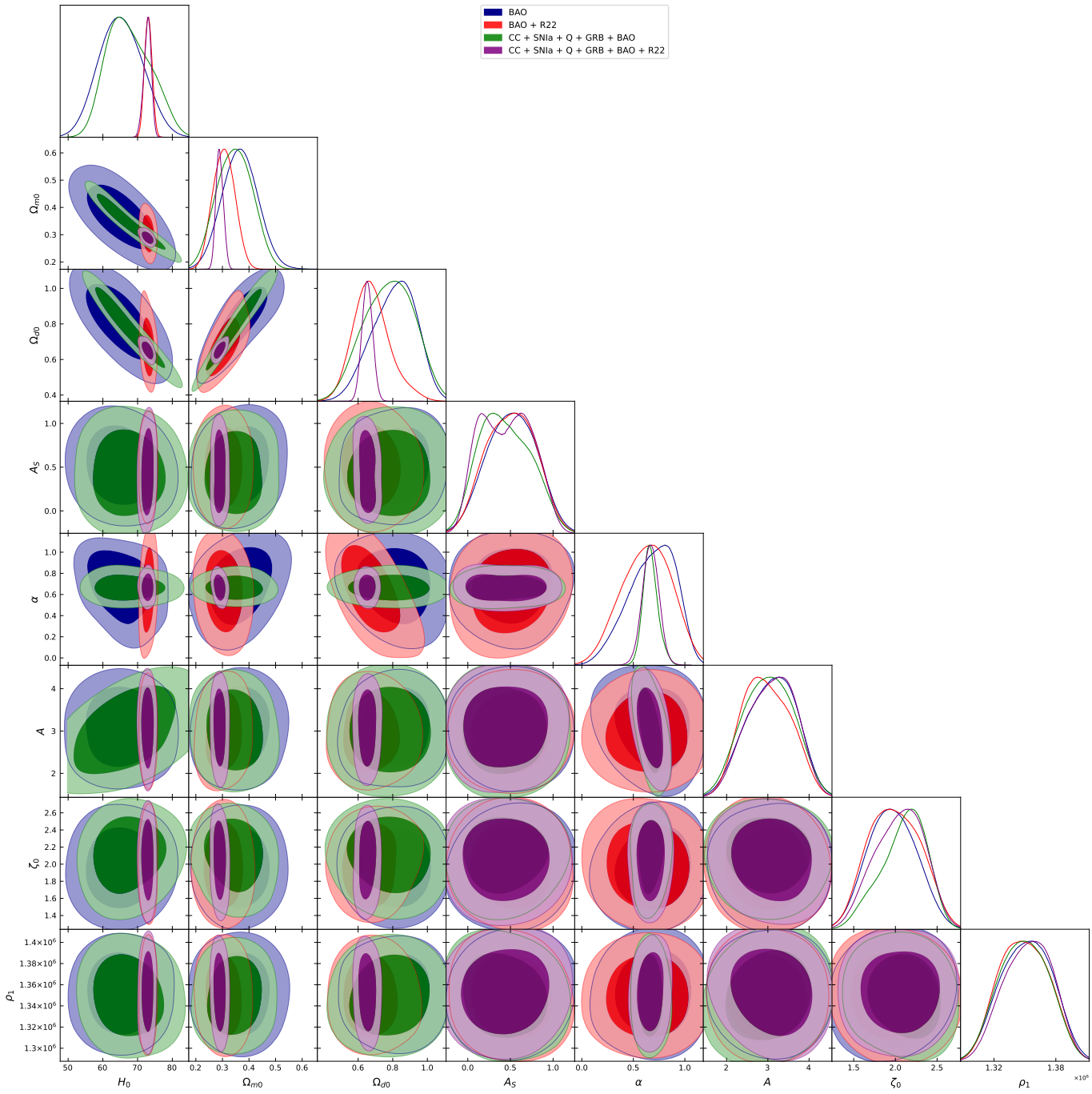


FIG. 3: The figure illustrates the posterior distribution of various observational data measurements using the Viscous Modified Chaplygin Gas model, highlighting the 1σ and 2σ regions.

to the H_0 tension. It is noteworthy that our results are contingent on the range of priors for r_d and H_0, influencing the estimated values in the $r_d - H_0$ contour plane. An interesting observation is that when we exclude the R22 prior, the results for H_0 and r_d tend to align with the Planck and SDSS results, mitigating the tension observed in the absence of this particular prior.

Now, in the study of the gravitational lensing optical depth, we use the analytical expression given by eqn. (25). After constraining the model parameters of the DE models as discussed above and finding their best-fit values, we plot the graphs of $\tau(z_s)/F$ against redshift z_s for both models. Figs. 6 and 7,

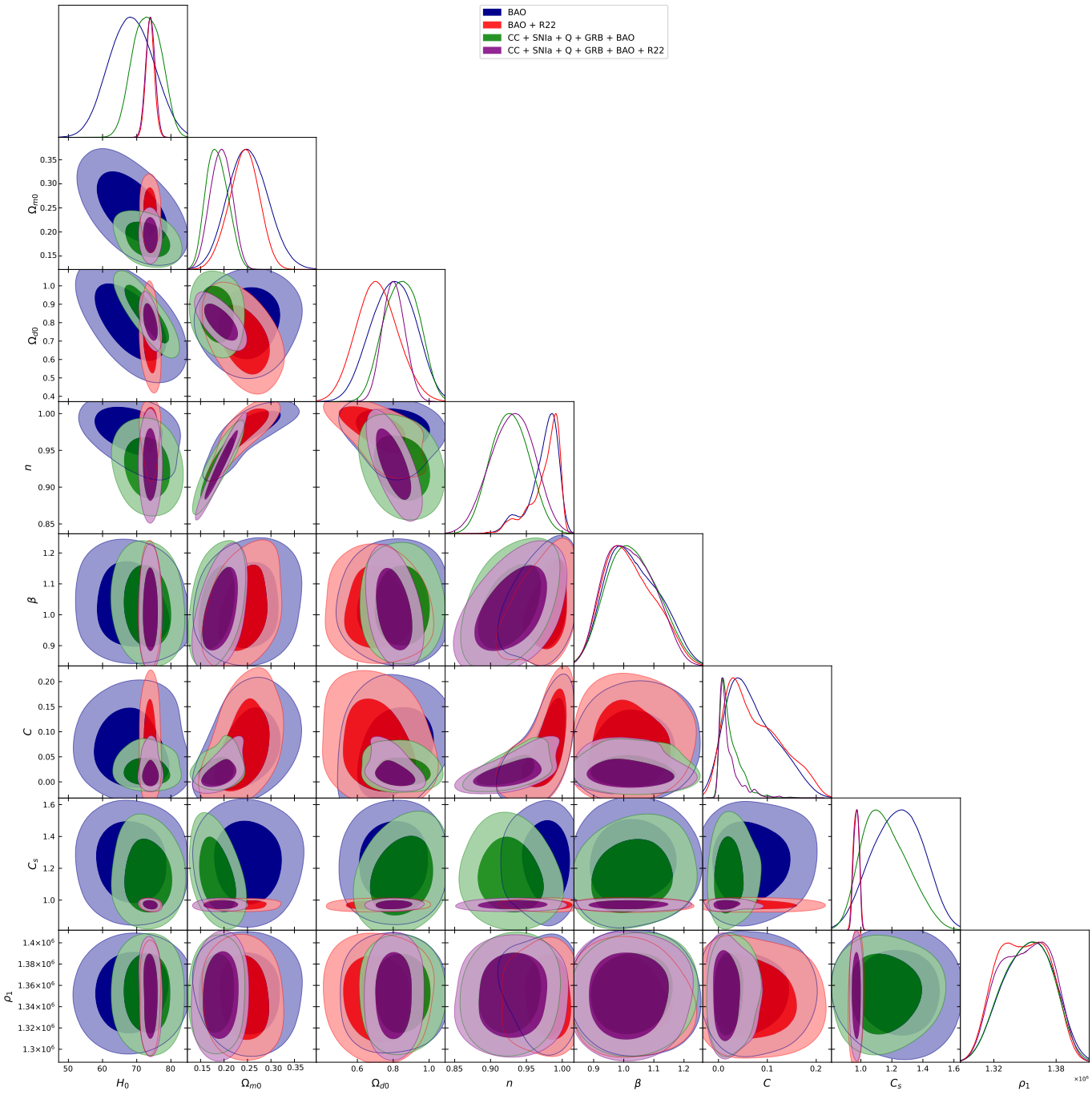
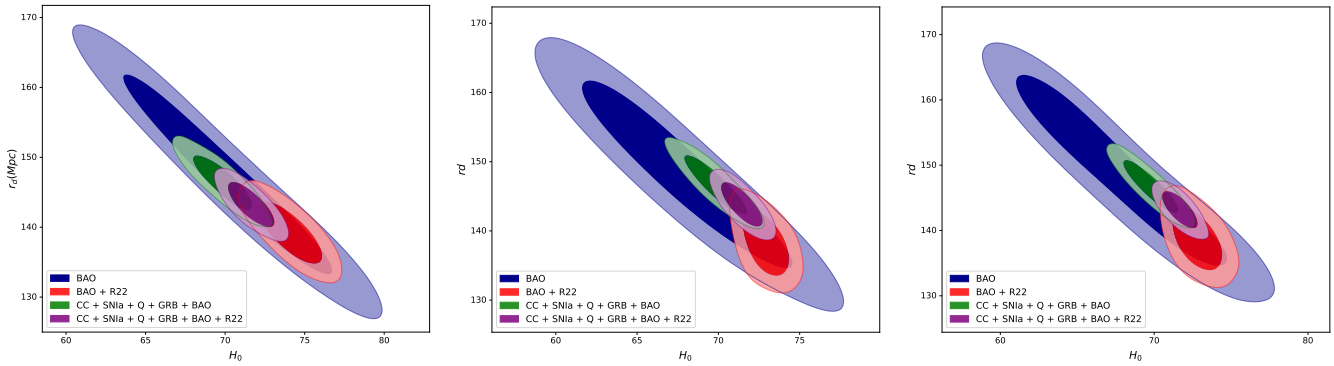


FIG. 4: The figure illustrates the posterior distribution of various observational data measurements using the Variable Modified Chaplygin Gas model, highlighting the 1σ and 2σ regions.

respectively, show changes in the optical depth behaviour of the two DE models here w.r.t z_s. We notice that the probability of finding gravitational lenses (i.e. optical depth) increases for both the models with increasing z_s value. Further, from fig. 6, we observe that the graphs of $\tau(z_s)$ for parameter sets constrained through different data sets show similar behaviour except for the parameters constraint through *BAO* only. However, for $z_s < 1$, they all coincide with each other. From these, we can conclude that the parameter values constraint through different data sets have very little to no effect in detecting the distribution of gravitational lenses in our Universe. Again, if we analyse fig. 7, we observe that parameters constrained



a) ΛCDM Model b) VsMCG Model c) VMCG Model

FIG. 5: The figure shows the posterior distribution of diverse observational data measurements within the $r_d - H_0$ contour plane using the ΛCDM, VsMCG and VMCG models. The shaded regions correspond to the 1σ and 2σ confidence plane.

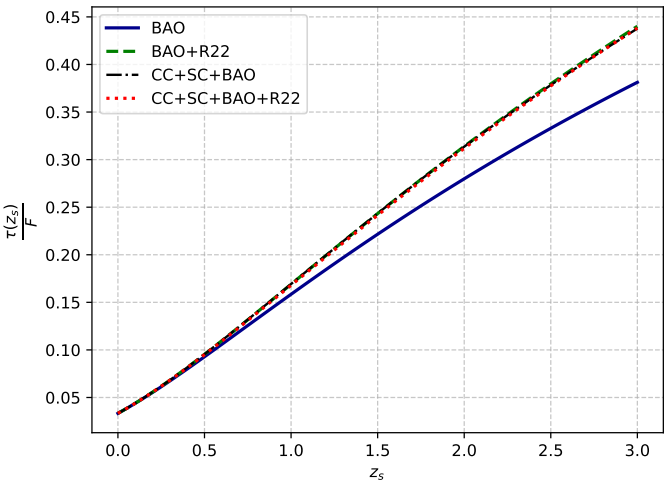


FIG. 6: Variation in optical depth behaviour of VsMCG model w.r.t z_s for parameters constrained through different data sets given in Table II.

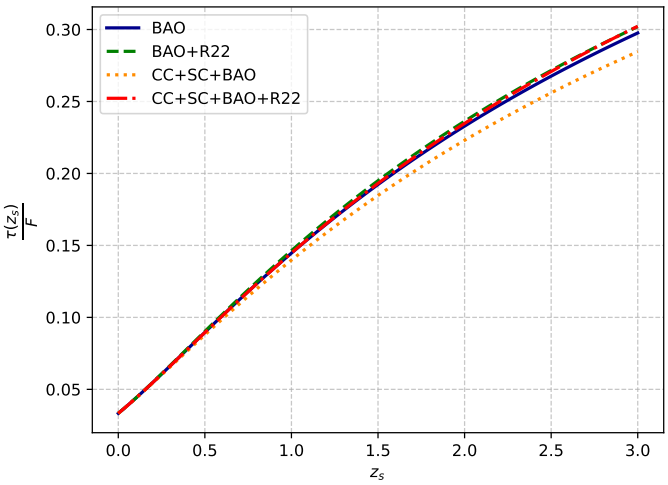


FIG. 7: Variation in optical depth behaviour of VMCG model w.r.t z_s for parameters constrained through different data sets given in Table II.

through the combination of $CC + SC + BAO + R22$ give a lower probability of finding gravitational lenses as compared to the other three combinations. However, for the other three combinations of data sets, viz. $BAO, BAO + R22, CC + SC + BAO$, we can observe small deviations in $\tau(z_s)$ for higher redshift values but for $z_s < 1$, the graph shows same lensing probability. This implies that in this DE model, parameter values constrained through $CC + SC + BAO + R22$ data sets give a slightly better approximation of the distribution of gravitational lenses than the rest. Finally, we conclude our study by plotting the optical depth of the two DE models considered here along with ΛCDM against source redshift z_s. We observe from Fig. 8 that although at the present epoch, the gravitational lensing probability for the three models coincides with each other, however moving back in time (or higher redshift values), all the models highly diverge from each other with ΛCDM gives the highest lensing probability and VMCG the lowest.

VI. DISCUSSIONS AND CONCLUSIONS

Assuming our Universe to be filled with dark matter and dark energy combinations in a flat FRW Universe, we reviewed two models of modified Chaplygin gas as the fluid source. For the dark energy candidates, we

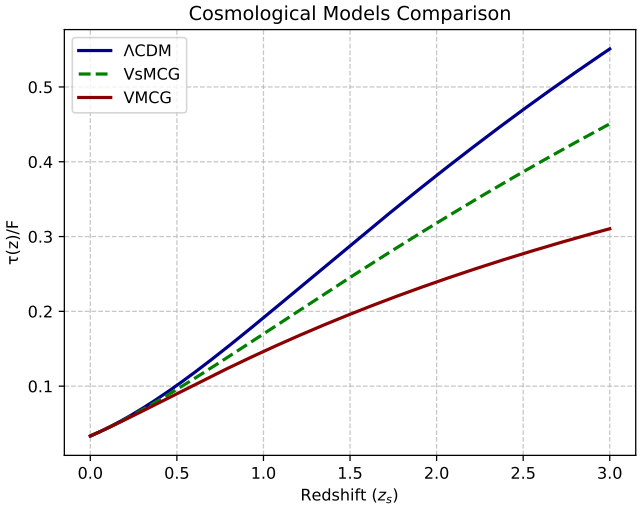


FIG. 8: Variation in optical depth behaviour of ΛCDM ($H_0 = 71.67, \Omega_{m0} = 0.26$), VsMCG ($H_0 = 72.86, \Omega_{m0} = 0.28, A_s = 0.44, \alpha = 0.66, A = 3.09, \zeta = 2.05, \rho_1 = 1.35351 \times 10^6$) and VMCG ($H_0 = 74.02, \Omega_{m0} = 0.18, n = 0.97, \beta = 1.02, C = 0.03, C_s = 0.48, \rho_1 = 1.35085 \times 10^6$) model w.r.t source redshift z_s.

considered viscous modified Chaplygin gas (VsMCG) and variable modified Chaplygin gas (VMCG) in the framework of loop quantum cosmology. We determined the respective Hubble parameter and, consequently, constraint the associated model parameters with the help of the latest observational data. Our investigation involved the selection of 24 Baryon Acoustic Oscillation (BAO) points were deliberately chosen to minimize correlations from an extensive dataset comprising 333 BAO data points. The objective was to mitigate inherent errors in the posterior distribution, which could potentially stem from correlations between measurements. Employing a method that mimics the introduction of random correlations into the covariance matrix, we conducted a systematic procedure. Our subsequent verification process attested that these introduced correlations do not lead to substantial alterations in the resulting cosmological parameters. Furthermore, we incorporated a variety of datasets into our analysis. This involved the inclusion of 30 uncorrelated data points from Cosmic Chronometers, a carefully selected set of 40 points obtained from Type Ia supernovae, 24 points extracted from the Hubble diagram for quasars, and a considerable dataset consisting of 162 points derived from Gamma Ray Bursts. Additionally, we integrated the most recent measurement of the Hubble constant conducted by researcher R22. Our recent investigation, conducted through a comprehensive array of observational assessments, brings attention to the persistent existence of the Hubble tension, although it is somewhat mitigated to a 2σ level for H_0. By introducing the sound horizon (r_d) as a free parameter, we have derived specific values for H_0 and r_d in various cosmological models, including the Standard ΛCDM, VsMCG, and VMCG models. In the ΛCDM model, our analysis reveals $H_0 = 69.828145 \pm 1.009964$ km/s/Mpc and $r_d = 146.826212 \pm 2.201612$ Mpc. For the VsMCG model, we obtain $H_0 = 67.447204 \pm 8.878032$ km/s/Mpc and $r_d = 147.559044 \pm 2.723020$ Mpc. Similarly, in the case of VMCG, the results indicate $H_0 = 69.204717 \pm 8.056031$ km/s/Mpc and $r_d = 147.906003 \pm 2.386071$ Mpc. Crucially, our assessments highlight that the values of H_0 and r_d based on low-redshift measurements align with early Planck estimates [15]. We also studied the gravitational lensing optical depth of our two dark energy models with the help of the constrained model parameter values. The graph showing changes in optical depth is plotted against their source redshift (fig. 6 & 7). For both models, the probability of finding gravitational lenses increases with higher redshift values. However, for low redshift ($z_s < 1$), the lensing probability in both the DE models coincides irrespective of the different sets of constrained parameter values obtained by considering different combinations of data sets. Lastly, we also carried out a joint analysis of the two DE models considered here with that of ΛCDM model (fig. 8). We found that the three models are indistinguishable for low redshift value ($z_s < 1$); however, they highly diverge from each other in the early Universe. Finally, we conclude our study with a statistical evaluation of our cosmological models and utilize both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The AIC is expressed as [65, 66]: $\text{AIC} = -2\ln(\mathcal{L}_{\max}) + 2k + \frac{2k(2k+1)}{N_{\text{tot}}-k-1}$ Here, $\mathcal{L}_{\max}$ signifies the maximum likelihood of the data, incorporating the entire dataset without the R22 prior. The parameters encompass N_{tot}, the total number of data points (in our instance, $N_{\text{tot}} = 296$), and k, the number of parameters. For large N_{tot}, this expression simplifies to:$\text{AIC} \simeq -2\ln(\mathcal{L}_{\max}) + 2k$ which is the conventional form of the AIC criterion [65]. In contrast, the Bayesian Information Criterion is formulated as [66]: $\text{BIC} = -2\ln(\mathcal{L}_{\max}) + k\ln N_{\text{tot}}$. By applying these criteria, we compute the AIC and BIC for the standard ΛCDM, VsMCG, and VMCG models. The obtained values for ΛCDM, VsMCG, and VMCG models are, respectively AIC = [277.38, 281.16, 281.12] and BIC = [277.59, 283.12, 282.34]. Despite the ΛCDM model displaying the best fit due to the lowest AIC, our collective AIC and BIC results lend support to all

the tested models. This suggests that none of the models can be dismissed based on the existing data. In the evaluation of the VsMCG and VMCG models relative to ΛCDM, we acknowledge that ΛCDM is embedded within both proposed extensions, differing by 5 degrees of freedom. This distinction allows for the application of standard statistical tests. The yardstick for comparison is the reduced chi-square statistic, defined as $\chi^2_{\text{red}} = \chi^2/\text{Dof}$, where Dof represents the degrees of freedom of the model, and χ^2 denotes the weighted sum of squared deviations with an equal number of runs for the three models, the statistic approximates 1, expressed as: $\left(\frac{\chi^2}{Dof_{\Lambda CDM}}, \frac{\chi^2}{Dof_{VsMCG}}, \frac{\chi^2}{Dof_{VMCG}}\right) \approx \{0.961646, 0.971234, 0.954310\}$. This comparative analysis offers valuable insights into the goodness of fit for each model, with values near 1 signifying a satisfactory alignment with the observed data. The findings suggest that both the VsMCG and VMCG models provide viable alternatives to the standard ΛCDM model, with comparable goodness of fit and support from statistical criteria.

[1] A. G. Riess, A. V. Filippenko, and et al., "Observational evidence from supernovae for an accelerating universe and a cosmological constant," *Astron. J.*, vol. 116, no. 3, p. 1009, 1998.

[2] D. N. Spergel, L. Verde, and et al., "First-year Wilkinson Microwave Anisotropy Probe (WMAP)* observations: determination of cosmological parameters," *Astrophys. J. Suppl. Ser.*, vol. 148, no. 1, p. 175, 2003.

[3] A. R. Liddle and D. H. Lyth, *Cosmological inflation and large-scale structure*. Cambridge university press, 2000.

[4] E. Komatsu, J. Dunkley, and et al., "FIVE-YEAR WILKINSON MICROWAVE ANISOTROPY PROBE OBSERVATIONS: COSMOLOGICAL INTERPRETATION," *Astrophys. J. Suppl. Ser.*, vol. 180, pp. 330–376, feb 2009.

[5] A. Kamenshchik, U. Moschella, and V. Pasquier, "An alternative to quintessence," *Physics Letters B*, vol. 511, no. 2-4, pp. 265–268, 2001.

[6] U. Debnath, A. Banerjee, and S. Chakraborty, "Role of modified chaplygin gas in accelerated universe," *Classical and Quantum Gravity*, vol. 21, no. 23, p. 5609, 2004.

[7] P. Peebles and B. Ratra, "Cosmology with a time-variable cosmological'constant'," *Astrophysical Journal, Part 2-Letters to the Editor (ISSN 0004-637X), vol. 325, Feb. 15, 1988, p. L17-L20. NSF-supported research.*, vol. 325, pp. L17–L20, 1988.

[8] R. R. Caldwell, R. Dave, and P. J. Steinhardt, "Cosmological imprint of an energy component with general equation of state," *Physical Review Letters*, vol. 80, no. 8, p. 1582, 1998.

[9] U. Alam, V. Sahni, T. Deep Saini, and A. A. Starobinsky, "Is there supernova evidence for dark energy metamorphosis?," *Monthly Notices of the Royal Astronomical Society*, vol. 354, no. 1, pp. 275–291, 2004.

[10] S. D. Hsu, "Entropy bounds and dark energy," *Phys. Lett. Sect. B Nucl. Elem. Part. High-Energy Phys.*, vol. 594, no. 1-2, pp. 13–16, 2004.

[11] M. Li, "A model of holographic dark energy," *Phys. Lett. Sect. B Nucl. Elem. Part. High-Energy Phys.*, vol. 603, no. 1-2, pp. 1–5, 2004.

[12] E. J. COPELAND, M. SAMI, and S. TSUJIKAWA, "DYNAMICS OF DARK ENERGY," *Int. J. Mod. Phys. D*, vol. 15, pp. 1753–1935, nov 2006.

[13] V. Springel, C. S. Frenk, and S. D. White, "The large-scale structure of the universe," *nature*, vol. 440, no. 7088, pp. 1137–1144, 2006.

[14] A. Jenkins, C. Frenk, F. Pearce, P. Thomas, J. Colberg, S. D. White, H. Couchman, J. Peacock, G. Efstathiou, and A. Nelson, "Evolution of structure in cold dark matter universes," *The Astrophysical Journal*, vol. 499, no. 1, p. 20, 1998.

[15] P. Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. Banday, R. Barreiro, N. Bartolo, *et al.*, "Planck 2018 results. vi. cosmological parameters," 2020.

[16] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, *et al.*, "Observational evidence from supernovae for an accelerating universe and a cosmological constant," *The astronomical journal*, vol. 116, no. 3, p. 1009, 1998.

[17] D. J. Eisenstein and W. Hu, "Baryonic features in the matter transfer function," *The Astrophysical Journal*, vol. 496, no. 2, p. 605, 1998.

[18] P. J. Peebles and J. Yu, "Primeval adiabatic perturbation in an expanding universe," *Astrophysical Journal, vol. 162, p. 815*, vol. 162, p. 815, 1970.

[19] D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H.-J. Seo, M. Tegmark, Z. Zheng, *et al.*, "Detection of the baryon acoustic peak in the large-scale correlation function of sdss luminous red galaxies," *The Astrophysical Journal*, vol. 633, no. 2, p. 560, 2005.

[20] S. Alam, F. D. Albareti, C. A. Prieto, F. Anders, S. F. Anderson, T. Anderton, B. H. Andrews, E. Armengaud, É. Aubourg, S. Bailey, *et al.*, "The eleventh and twelfth data releases of the sloan digital sky survey: final data from sdss-iii," *The Astrophysical Journal Supplement Series*, vol. 219, no. 1, p. 12, 2015.

[21] S. NOJIRI and S. D. ODINTSOV, "INTRODUCTION TO MODIFIED GRAVITY AND GRAVITATIONAL ALTERNATIVE FOR DARK ENERGY," *Int. J. Geom. Methods Mod. Phys.*, vol. 04, pp. 115–145, feb 2007.

[22] S. Capozziello and M. De Laurentis, "Extended Theories of Gravity," *Phys. Rep.*, vol. 509, pp. 167–321, dec 2011.

[23] A. Ashtekar and J. Lewandowski, "Background independent quantum gravity: a status report," *Classical and Quantum Gravity*, vol. 21, no. 15, p. R53, 2004.

[24] M. Bojowald, "Loop quantum cosmology," *Living Reviews in Relativity*, vol. 11, pp. 1–131, 2008.

[25] M. Jamil and U. Debnath, "Interacting modified chaplygin gas in loop quantum cosmology," *Astrophysics and Space Science*, vol. 333, pp. 3–8, 2011.

[26] X. Fu, H. Yu, and P. Wu, "Dynamics of interacting phantom scalar field dark energy in loop quantum cosmology," *Physical Review D*, vol. 78, no. 6, p. 063001, 2008.

[27] S. Refsdal and H. Bondi, "The gravitational lens effect," *Monthly Notices of the Royal Astronomical Society*, vol. 128, no. 4, pp. 295–306, 1964.

[28] E. L. Turner, J. P. Ostriker, and J. R. Gott III, "The statistics of gravitational lenses-the distributions of image angular separations and lens redshifts," *The Astrophysical Journal*, vol. 284, pp. 1–22, 1984.

[29] J. R. Gott III, M.-G. Park, and H. M. Lee, "Setting limits on q0 from gravitational lensing," *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 338, March 1, 1989, p. 1-12.*, vol. 338, pp. 1–12, 1989.

[30] Z.-H. Zhu, "Gravitational lensing statistical properties in general frw cosmologies with dark energy component (s): analytic results," *International Journal of Modern Physics D*, vol. 9, no. 05, pp. 591–600, 2000.

[31] Z.-H. Zhu, "Gravitational lensing statistics as a probe of dark energy," *Modern Physics Letters A*, vol. 15, no. 16, pp. 1023–1029, 2000.

[32] R. Freitas, S. Gonçalves, and A. Oliveira, "Gravitational lenses in the dark universe," *Astrophysics and Space Science*, vol. 349, pp. 443–455, 2014.

[33] R. Kundu, U. Debnath, and A. Pradhan, "Gravitational lensing: dark energy models in non-flat frw universe," *The European Physical Journal C*, vol. 83, no. 6, p. 553, 2023.

[34] D. Aberkane, N. Mebarki, and S. Benchikh, "Viscous modified chaplygin gas in classical and loop quantum cosmology," *Chinese Physics Letters*, vol. 34, no. 6, p. 069801, 2017.

[35] U. Debnath, "Variable modified chaplygin gas and accelerating universe," *Astrophysics and Space Science*, vol. 312, no. 3-4, pp. 295–299, 2007.

[36] A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden, and M. Manera, "The clustering of the sdss dr7 main galaxy sample–i. a 4 per cent distance measure at z= 0.15," *Monthly Notices of the Royal Astronomical Society*, vol. 449, no. 1, pp. 835–847, 2015.

[37] S. Alam, M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek, A. S. Bolton, J. R. Brownstein, A. Burden, C.-H. Chuang, *et al.*, "The clustering of galaxies in the completed sdss-iii baryon oscillation spectroscopic survey: cosmological analysis of the dr12 galaxy sample," *Monthly Notices of the Royal Astronomical Society*, vol. 470, no. 3, pp. 2617–2652, 2017.

[38] H. Gil-Marín, J. E. Bautista, R. Paviot, M. Vargas-Magaña, S. de La Torre, S. Fromenteau, S. Alam, S. Ávila, E. Burtin, C.-H. Chuang, *et al.*, "The completed sdss-iv extended baryon oscillation spectroscopic survey: measurement of the bao and growth rate of structure of the luminous red galaxy sample from the anisotropic power spectrum between redshifts 0.6 and 1.0," *Monthly Notices of the Royal Astronomical Society*, vol. 498, no. 2, pp. 2492–2531, 2020.

[39] A. Raichoor, A. De Mattia, A. J. Ross, C. Zhao, S. Alam, S. Avila, J. Bautista, J. Brinkmann, J. R. Brownstein, E. Burtin, *et al.*, "The completed sdss-iv extended baryon oscillation spectroscopic survey: large-scale structure catalogues and measurement of the isotropic bao between redshift 0.6 and 1.1 for the emission line galaxy sample," *Monthly Notices of the Royal Astronomical Society*, vol. 500, no. 3, pp. 3254–3274, 2021.

[40] J. Hou, A. G. Sánchez, A. J. Ross, A. Smith, R. Neveux, J. Bautista, E. Burtin, C. Zhao, R. Scoccimarro, K. S. Dawson, *et al.*, "The completed sdss-iv extended baryon oscillation spectroscopic survey: Bao and rsd measurements from anisotropic clustering analysis of the quasar sample in configuration space between redshift 0.8 and 2.2," *Monthly Notices of the Royal Astronomical Society*, vol. 500, no. 1, pp. 1201–1221, 2021.

[41] H. D. M. Des Bourboux, J. Rich, A. Font-Ribera, V. de Sainte Agathe, J. Farr, T. Etourneau, J.-M. Le Goff, A. Cuceu, C. Balland, J. E. Bautista, *et al.*, "The completed sdss-iv extended baryon oscillation spectroscopic survey: baryon acoustic oscillations with lyα forests," *The Astrophysical Journal*, vol. 901, no. 2, p. 153, 2020.

[42] T. Abbott, M. Aguena, S. Allam, A. Amon, F. Andrade-Oliveira, J. Asorey, S. Avila, G. Bernstein, E. Bertin, A. Brandao-Souza, *et al.*, "Dark energy survey year 3 results: A 2.7% measurement of baryon acoustic oscillation distance scale at redshift 0.835," *Physical Review D*, vol. 105, no. 4, p. 043512, 2022.

[43] S. Sridhar, Y.-S. Song, A. J. Ross, R. Zhou, J. A. Newman, C.-H. Chuang, R. Blum, E. Gaztanaga, M. Landriau, and F. Prada, "Clustering of lrgs in the decals dr8 footprint: Distance constraints from baryon acoustic oscillations using photometric redshifts," *The Astrophysical Journal*, vol. 904, no. 1, p. 69, 2020.

[44] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson, "The 6df galaxy survey: baryon acoustic oscillations and the local hubble constant," *Monthly Notices of the Royal Astronomical Society*, vol. 416, no. 4, pp. 3017–3032, 2011.

[45] L. Kazantzidis and L. Perivolaropoulos, "Evolution of the fσ 8 tension with the planck 15/λcdm determination and implications for modified gravity theories," *Physical Review D*, vol. 97, no. 10, p. 103503, 2018.

[46] M. Moresco, "Raising the bar: new constraints on the hubble parameter with cosmic chronometers at z 2," *Monthly Notices of the Royal Astronomical Society: Letters*, vol. 450, no. 1, pp. L16–L20, 2015.

[47] M. Moresco, L. Pozzetti, A. Cimatti, R. Jimenez, C. Maraston, L. Verde, D. Thomas, A. Citro, R. Tojeiro, and D. Wilkinson, "A 6% measurement of the hubble parameter at z 0.45: direct evidence of the epoch of cos-

mic re-acceleration," *Journal of Cosmology and Astroparticle Physics*, vol. 2016, no. 05, pp. 014–014, 2016.

[48] M. Moresco, L. Verde, L. Pozzetti, R. Jimenez, and A. Cimatti, "New constraints on cosmological parameters and neutrino properties using the expansion rate of the universe to z 1.75," *Journal of Cosmology and Astroparticle Physics*, vol. 2012, no. 07, p. 053, 2012.

[49] M. Moresco, A. Cimatti, R. Jimenez, L. Pozzetti, G. Zamorani, M. Bolzonella, J. Dunlop, F. Lamareille, M. Mignoli, H. Pearce, *et al.*, "Improved constraints on the expansion rate of the universe up to z 1.1 from the spectroscopic evolution of cosmic chronometers," *Journal of Cosmology and Astroparticle Physics*, vol. 2012, no. 08, pp. 006–006, 2012.

[50] D. M. Scolnic, D. Jones, A. Rest, Y. Pan, R. Chornock, R. Foley, M. Huber, R. Kessler, G. Narayan, A. Riess, *et al.*, "The complete light-curve sample of spectroscopically confirmed sne ia from pan-starrs1 and cosmological constraints from the combined pantheon sample," *The Astrophysical Journal*, vol. 859, no. 2, p. 101, 2018.

[51] C. Roberts, K. Horne, A. O. Hodson, and A. D. Leggat, "Tests of λcdm and conformal gravity using grb and quasars as standard candles out to $z \sim 8$," *arXiv preprint arXiv:1711.10369*, 2017.

[52] M. Demianski, E. Piedipalumbo, D. Sawant, and L. Amati, "Cosmology with gamma-ray bursts-ii. cosmography challenges and cosmological scenarios for the accelerated universe," *Astronomy & Astrophysics*, vol. 598, p. A113, 2017.

[53] A. G. Riess, W. Yuan, L. M. Macri, D. Scolnic, D. Brout, S. Casertano, D. O. Jones, Y. Murakami, G. S. Anand, L. Breuval, *et al.*, "A comprehensive measurement of the local value of the hubble constant with 1 km s- 1 mpc- 1 uncertainty from the hubble space telescope and the sh0es team," *The Astrophysical journal letters*, vol. 934, no. 1, p. L7, 2022.

[54] W. Handley, M. Hobson, and A. Lasenby, "Polychord: nested sampling for cosmology," *Monthly Notices of the Royal Astronomical Society: Letters*, vol. 450, no. 1, pp. L61–L65, 2015.

[55] A. Lewis, "Getdist: a python package for analysing monte carlo samples," *arXiv preprint arXiv:1910.13970*, 2019.

[56] J. R. Gott III and J. E. Gunn, "The double quasar 1548+ 115a, b as a gravitational lens," *Astrophysical Journal, vol. 190, p. L105*, vol. 190, p. L105, 1974.

[57] E. L. Turner, "Gravitational lensing limits on the cosmological constant in a flat universe," *Astrophys. J.*, vol. 365, pp. L43—-L46, 1990.

[58] M. Fukugita, T. Futamase, and M. Kasai, "A possible test for the cosmological constant with gravitational lenses," *Mon. Not. R. Astron. Soc.*, vol. 246, p. 24P, 1990.

[59] L. M. Krauss and M. White, "Gravitational lensing, finite galaxy cores, and the cosmological constant," *Astrophys. J.*, vol. 394, pp. 385–395, 1992.

[60] C. S. Kochanek, "Analytic results for the gravitational lens statistics of singular isothermal spheres in general cosmologies," *Mon. Not. R. Astron. Soc.*, vol. 261, no. 2, pp. 453–463, 1993.

[61] R. C. Nunes, S. K. Yadav, J. Jesus, and A. Bernui, "Cosmological parameter analyses using transversal bao data," *Monthly Notices of the Royal Astronomical Society*, vol. 497, no. 2, pp. 2133–2141, 2020.

[62] R. C. Nunes and A. Bernui, "Bao signatures in the 2-point angular correlations and the hubble tension," *The European Physical Journal C*, vol. 80, pp. 1–8, 2020.

[63] L. Verde, J. L. Bernal, A. F. Heavens, and R. Jimenez, "The length of the low-redshift standard ruler," *Monthly Notices of the Royal Astronomical Society*, vol. 467, no. 1, pp. 731–736, 2017.

[64] T. Lemos, Ruchika, J. C. Carvalho, and J. Alcaniz, "Low-redshift estimates of the absolute scale of baryon acoustic oscillations," *The European Physical Journal C*, vol. 83, no. 6, p. 495, 2023.

[65] H. Akaike, "A new look at the statistical model identification," *IEEE transactions on automatic control*, vol. 19, no. 6, pp. 716–723, 1974.

[66] G. Schwarz, "Estimating the dimension of a model," *The annals of statistics*, pp. 461–464, 1978.

Sādhanā (2024) 49:75
https://doi.org/10.1007/s12046-023-02425-4
© Indian Academy of Sciences

Greenfield settlements due to tunnelling using tunnel boring machine (TBM) in layered soils: a parametric study

JYOTI AGARWAL* and RAJU SARKAR

Civil Engineering Department, Delhi Technological University, Shahbad Daulatpur Main Bawana Road, Delhi 110042, India
e-mail: jyotiagarwal_2k20phdce04@dtu.ac.in; rajusarkar@dce.ac.in

MS received 27 April 2023; revised 20 October 2023; accepted 22 November 2023

Abstract. As India's cities continue to increase in population, a large number of new metro lines are being built to accommodate the mobility of residents. For construction of these metro lines, underground tunnels are being constructed to alleviate traffic problems in congested areas such as Delhi, Chennai, Mumbai, etc., where land scarcity is a major issue. Although tunnel boring machines (TBMs) are used in the construction of metro tunnels, ground settlements on the surface are inevitable. Tunnelling induced ground settlements may have adverse impact on the structures that are above the ground level as well as adjacent to the excavations. In the present study, ground settlements caused by TBM tunnelling are determined in greenfield conditions, using a finite element method (FEM) based software, PLAXIS and parametric study is performed after validation of numerical model with empirical results. Ground settlements obtained using PLAXIS 2D are compared with those obtained using empirical methods as available in literature. This study captures the effect of governing tunnel and soil parameters such as diameter of tunnel (D), depth of the tunnel axis from ground surface (z), grout pressure (GP), undrained shear strength of soil (s_u) and soil modulus (E'_{ref}), covering the most practical range, on ground settlements under greenfield condition. The settlement results of this study are validated and compared closely with earlier experimental and observational findings that have been published in the literature. It is observed from the various parameters studied that they are having a major impact on the ground settlements during tunnel construction. Therefore, a tunnel designer should consider the effect of all the governing factors in design prior to the tunnel construction to keep the maximum surface settlements within the permissible limit set by the standards.

Keywords. Tunnelling; greenfield settlements; volume loss; parametric study.

1. Introduction

Research in the transportation services for better management and infrastructure development has been prompted by population growth, urbanisation, and the quick expansion of emerging economies. Delhi and Mumbai are predicted to rank among the world's two most populous cities by the year 2035. Rapid transit system advancements have followed the rise in land scarcity, urbanisation, and industrialisation caused by the rapid population growth rate [1]. With little room for alteration above ground, tunnels have become a more popular option, making greater use of the underground area is a good alternative. Tunnelling induced ground settlement is a complex phenomenon that depends on a variety of factors including the soil and tunnel parameters.

Many researchers have performed centrifuge model tests in clayey and sandy soils to estimate ground deformations caused by tunnelling and compared recorded ground movements with results from existing empirical and analytical approaches [2]. Loganathan *et al* [2] evaluated the tunnelling-induced ground deformations and their impact on adjacent pile foundations by conducting three centrifuge model tests in clays. Jacobsz *et al* [3] carried out a study using centrifuge to examine how tunnelling in dense, dry sand will affect surrounding single piles and found that volume loss resulted by tunnelling may cause large pile settlements within a zone of influence. Marshall *et al* [4] analyse ground movement by conducting centrifuge tests due to tunnelling in sands and assessed the impact of tunnel dimensions and volume loss on soil deformation under greenfield conditions. Many researchers have given analytical methods to anticipate tunnelling-induced ground movements in clayey and sandy soils [5, 6]. Verruijt and Booker [6] proposed an analytical solution for determining the uniform radial displacement due to ground loss and ovalization of tunnel during tunnelling in an elastic soil. Loganathan and Poulos [7] proposed an analytical solution

*For correspondence
Published online: 15 February 2024

by redefining the term ground loss as an equivalent ground loss parameter with regard to gap parameter in order to predict ground deformations due to tunnelling in undrained condition, and results of this study are found to be satisfactory especially for the tunnels in stiff clays. Vorster *et al* [5] presented an analytical method for calculating the maximum bending moment for pipelines that are rigidly joined or continuous and affected by ground movement resulting from tunnelling. The proposed approach was found conservative as compared with centrifuge tests. Fu *et al* [8] proposed an analytical solution to predict ground movement caused by a circular tunnel that is nonuniformly deforming in a heavy elastic half-plane. The proposed solution was found to be very useful in the preliminary design of tunnels. Lai *et al* [9] proposed a new approach to determine the tunnelling-induced ground settlements by utilizing the centrifuge model test results for loess soils. Hong *et al* [10] proposed an analytical model based on complex variable theory to investigate ground responses due to shallow tunnelling in multi-layered ground with an arbitrary ground surface load. Based on the principle of linear-elastic superposition, the multi-boundary problem is converted into a superposition of multiple single-boundary problems. The accuracy of the proposed method was validated using the field data and numerical analysis results. Sun *et al* [11] predicted the tunnel settlement in structural soft soils during construction by numerically simulated disturbance degree calculated based on effective stress and developed a two-fold line settlement calculation model by combining it with one-dimensional consolidation test results for disturbed soil.

As computer technology has advanced, numerical modelling utilising finite element analysis has become the dominant method for geotechnical design and analysis. Results of numerical models can be compared to empirical and semi-empirical methods that are nonetheless useful and capable. Numerous academics have already concentrated on surface settlement prediction under greenfield conditions due to tunnelling by employing numerical modelling techniques [1, 12–22]. Möller and Vermeer [12] performed a 2D finite element analysis to assess lining forces and deformations caused by closed shield and open face tunnelling. Mathew and Lehane [13, 14] used PLAXIS 2D and PLAXIS 3D to do a numerical back analysis of greenfield deformations caused by twin bored tunnels. Three soil constitutive models were used, and parametric study was performed. Zheng *et al* [15] proposed a fuzzy statistics-based method for the calculation of the standard values and deviations of mechanical driving parameters. In order to study the impact of the mechanical driving parameters on ground movement, a finite element method was used. Golpasand *et al* [16] calculated the real volume loss using semiempirical, numerical methods via performing back analysis on real settlements. Numerical modelling was performed using FLAC 3D code. Miliziano and Lillis [17] developed a PLAXIS 3D model to examine the soil, tunnel and building interaction along with the effects of mechanized tunnelling on the ground surface during the construction of Rome's metro line C. Shiau and Sams [18] developed a numerical method for assessing the settlements associated with circular tunnels in soft soils and presented design charts using dimensionless ratios with respect to geometry, strength and stiffness parameters. Settlement results of this study compare closely with earlier experimental and observational findings. Kanagaraju and Krishnamurthy [19] presented the profiling of influence zone in cohesionless deposits using a finite element programme PLAXIS 3D by analysing the tunnelling-induced settlements under free field or Greenfield conditions. Rezaei and Adli [20] performed a 3D finite element analysis using ABAQUS for simulation of Tabriz Metro Tunnel mechanized tunnelling and longitudinal troughs are obtained. Athar *et al* [1] performed physical and numerical modelling to examine the effect of lining shape on ground settlement. It was found that square linings are the least stable, whereas elliptical linings are best for minimising ground settlement. Aswathy *et al* [21] performed a numerical study using PLAXIS 3D and captured the impact of vital parameters such as grout pressure, depth and diameter of tunnel, etc., on surface settlement caused by tunnelling in young alluvium deposit. Ahmed *et al* [22] performed finite element analysis to simulate TBM construction processes using PLAXIS 3D and estimated the tunnelling-induced surface settlement for Metro Rail Line 1 of Dhaka.

1.1 *Ground settlements due to volume loss*

Ground settlements associated with tunnelling using modern tunnel boring machines (TBM's) are mainly due to the volume loss (V_L). When a tunnel is constructed, the soil surrounding the tunnel releases stress because of the delayed tunnel peripheral lining installation. As a result, radial, and longitudinal deformations occur in the soil body that tends the ground surrounding the tunnel moves into the tunnel cavity [8, 23, 24]. Volume loss (V_L) is represented as volume of ground loss around a tunnel cavity (V_T) with respect to the tunnel volume per unit length of tunnel [2, 7, 18, 25–27]. Total volume loss includes ground loss at the face of the tunnel (caused by movement of soil towards the excavation chamber), ground loss along the shield (due to progressive decrement of shield diameter from the cutter head to the TBM tail), ground loss at the tail (Due to the cavity between the tunnel lining and surrounding soil and movement of soil into the gap caused by reduction in grout pressure), ground loss behind the shield tail (caused by consolidation of grout) and long term volume loss caused by grout shrinkage and lining deformations [16, 23, 27–29]. Volume loss results in ground settlements. In case of clay considering non-dilation and undrained condition volume of surface settlement trough per unit length (V_s) should be

equal to the volume of soil moving into the tunnel cavity (V_T) [4, 18, 27]. Cohesionless soils like fills, silts, sands, and unsaturated clays provide some drainage which causes actual surface settlement trough to have a different volume than the volume of ground that moves into the tunnel cavity as a result of volume change within the soil [24]. Consolidation of clays may also cause sublayer subsidence resulting in additional ground movements along with those obtained due to volume loss during tunnelling activity [30]. The volume of the surface settlement trough may be less than the amount of ground that moves into the tunnel due to dilatation and swelling brought on by unloading that may cause soil expansion [12].

The survey of the literature reveals that several academics have worked extensively to estimate the ground deformations brought on by TBM tunnelling. However, the majority of the aforementioned research has been conducted outside of India. Large-scale TBM tunnel construction is currently underway in Delhi, India, in order to establish metro lines. Therefore, a thorough investigation is needed to evaluate the impact of TBM tunnelling on the ground of Indian region. In this paper, ground settlements due to TBM tunnelling are obtained using Finite Element Program PLAXIS 2D for Delhi, capital of India. To do this, input parameters for soil properties and tunnelling dimensions are used from one of the tunnelling projects ongoing in Delhi national capital region. Results of finite element analysis are validated using empirical methods. Finally, a parametric study has been conducted to evaluate the effects of tunnel and soil parameters on ground settlements under greenfield condition.

2. Problem statement

The Delhi–Meerut Regional Rapid Transit System (Delhi–Meerut RRTS) is an 82.15 km long, semi-high speed rail corridor currently under construction which will connect Delhi, Ghaziabad, and Meerut. This corridor has two underground stretch, one 5.6 km New Ashok Nagar—Sahibabad underground package 4 (1 underground station at Anand Vihar) and other 5.68 km Brahampuri DN Ramp - Begumpul UP Ramp underground package 8 (3 underground stations at Meerut Central, Bhaisali and Begumpul). RRTS underground package 4 stretch will be done by 4 TBMs named as Sudarshan. TBM Sudarshan 4.1 and 4.2 will build 2.8 km tunnel from Anand Vihar launch shaft to Khichripur retrieval shaft located at the northern end of Ashok Nagar DN ramp. TBM Sudarshan 4.3 and 4.4 will build 2.09 km tunnels between Anand Vihar launch shaft and Sahibabad UP ramp [31]. For the present study, tunnel alignment for the stretch from Anand Vihar launch shaft to Khichripur retrieval shaft is considered for evaluation of tunnelling-induced ground settlements under greenfield conditions. Most of the relevant data required for the present work has been obtained using open-sourced tender documents [32]. Tunnel Alignment considered in this study is represented in figure 1.

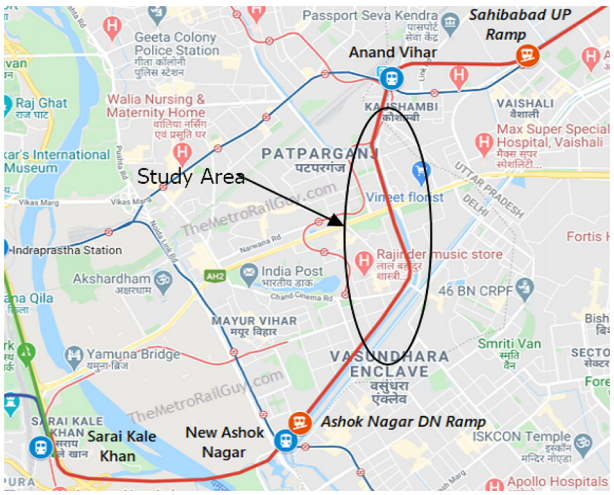


Figure 1. Tunnel alignment and the study area [31].

3. Soil and tunnel parameters for numerical modelling

Using field and laboratory test results given in Geotechnical investigation report, available in tender documents [32] interpretation and derivation of geological sub-strata profiles with suggested design parameters have been done for the stretch from Anand Vihar launch shaft to Khichripur retrieval shaft. The geology of the subject alignment mainly consists of two types of strata, fine silty sand (SM) and low plastic sandy silt (CL). Maximum water table was observed to be at 15 m below the ground level, as per the geotechnical investigation report. The ground profile is made up of eight layers that are distinguished by their thickness, soil type, physical and geotechnical properties. Physical properties and shear strength parameters of soils are taken from the drained and undrained laboratory test results given in Geotechnical investigation report, available in tender documents [32]. Soil modulus is obtained using Standard penetration test resistance, N and CIRIA report 143 by Clayton [33]. Table 1 summarizes the layer-wise physical and geotechnical properties of soil used for numerical analysis.

Tunnel is having an internal diameter of 6.5 m which is bigger than the other metro tunnels due to the bigger rolling stock and a high design speed of 180 kmph. Depth of Tunnel axis is 20 m below the ground surface. Tunnel lining is assumed to be made of 300 mm thick M50 grade concrete segments. Table 2 shows material properties of TBM and lining.

Table 1. Soil properties for numerical analysis [32–34].

Soil layer and parameter	Layer-1	Layer-2	Layer-3	Layer-4	Layer-5	Layer-6	Later-7	Layer-8
Layer type	SM	SM	SM	CL	CL	SM	SM	SM
Drainage type	Drained	Drained	Drained	Undrained B	Undrained B	Drained	Drained	Drained
Depth (m)	0–5	5–10	10–15	15–20	20–25	25–30	30–40	Below 40
Unsaturated unit weight, Υ_{unsat} (kN/m^3)	17	17	18	19	19	20	20	20
Saturated unit weight, Υ_{sat} (kN/m^3)	18	18	19	20	20	21	21	21
Young's modulus, E'$_{ref}$ (kPa)	8000	12000	20000	20000	29000	45000	50000	60000
Poisson's ratio, ν (nu)	0.38	0.38	0.38	0.35	0.35	0.38	0.38	0.38
Friction angle, φ' (phi)(°)	30	30	32	–	–	34	35	35
Cohesion, c'$_{ref}$ (kPa)	0	0	0	–	–	0	0	0
Undrained shear strength, s_u (kPa)	–	–	–	90	130	–	–	–

Table 2. Material properties of TBM and lining [35, 36].

Material parameters	Lining	TBM
Material type	Elastic	Elastic
Unit weight, w (kN/m^2)	7.5	19.5
Axial stiffness, EA$_1$ (kN/m)	10.6*10^6	50*10^6
Bending stiffness, EI (kN-m^2/m)	79.5*10^3	260*10^3
Poisson's ratio, ν (nu)	0.15	0.3

4. Finite element analysis

Even though tunnelling is a three-dimensional (3D) operation, transverse settlement under greenfield conditions can be estimated rather adequately in two-dimensional (2D) simulation. Numerous parameters are required for 3D numerical modelling, which makes it complicated, time consuming and computationally demanding. Moreover, parameters required for 3D modelling are sometimes challenging to determine in practice. For the sake of simplicity, it is reasonable to assume that the tunnel is very long and at a consistent depth [18].

4.1 *Numerical modelling*

In this study, numerical simulations have been performed using the finite element program PLAXIS 2D, 2022 to estimate the tunnelling-induced transverse settlements.

A plain strain 15-noded element model is used for analysis. In order to minimise the boundary effects the dimensions of the model have been considered 10 times the diameter of Tunnel in all directions. The model has been provided with standard fixities [35, 36]. Model consists of 650 elements with 5401 nodes after meshing. Final Model geometry and meshing used for FE analysis is shown in figures 2 and 3, respectively. Elastoplastic soil constitutive Mohr–Coulomb model, is used for modelling of soil layers in the current work. Table 1 summarizes soil properties

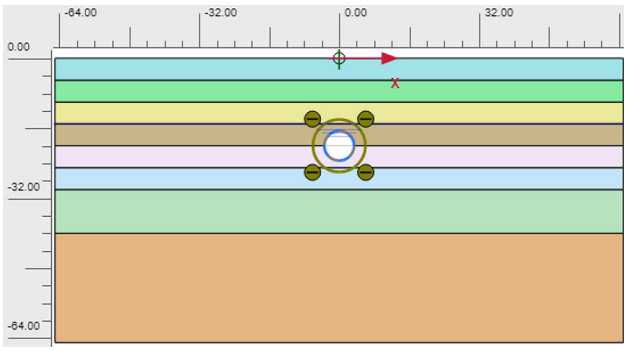


Figure 2. PLAXIS 2D model of existing ground conditions with tunnel.

used for numerical analysis. Table 2 summarizes material properties for TBM, and lining used for numerical analysis. Plate elements are used to model TBM and tunnel lining. For simulating the interaction between ground and TBM as well as ground and linings interface elements with strength reduction factor (R_{inter} = 0.67) for all the soil layers are provided. The water table is assigned to produce the hydrostatic initial pore water pressure and phreatic analysis has been done with the assumption of no groundwater flow during excavation and lining is impermeable. In the proposed study volume loss at tunnel (V_L) is taken as 1.5%.

The process of tunnel construction has been modelled through step-by-step method in 5 phases consists of: (1) Initial phase includes setting up of initial conditions (effective stresses and pore pressures) for the model before excavation with K0 procedure calculation type, (2) TBM phase includes the excavation of soil and activation of TBM, (3) TBM Conicity phase includes contraction of tunnel lining simulation due to the volume loss resulted from the cone shape of TBM, (4) Tail void grouting phase includes the simulation of application of grout pressure at the tail of the TBM. The grouting process is simulated by applying a pressure on the surrounding soil, (5) Lining installation phase includes the activation of final lining.

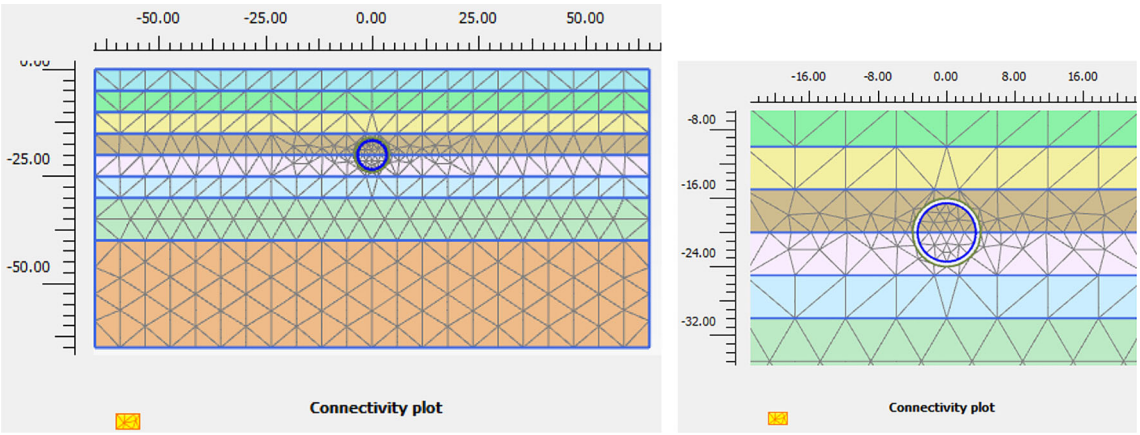


Figure 3. PLAXIS 2D Model after meshing.

Deformed mesh and Total vertical deformations after calculations are shown in figures 4 and 5, respectively.

4.2 *Validation*

Prior to assessing the influence of governing parameters on the ground settlements induced by tunnelling, the greenfield surface settlement trough obtained by numerical modelling has been compared with the surface settlement trough from empirical methods proposed in literature. Figure 6 shows the PLAXIS output for the vertical settlements at the ground surface. Here, validation of finite element model is done with four empirical relations given in literature.

Empirical method proposed by Peck *et al* [26] and Attewell and Farmer [27] is most widely used for calculating tunnelling induced ground settlements because of its suitability and ease of use. This method typically follows a Gaussian distribution curve when predicting ground settlements. The primary reasons for its adaptability are its shape which shows a close fit with the commonly observed experimental and field results, and its requirement of few easy to determine parameters [7]. This method calculates ground surface settlement (S_x) at a horizontal distance (x)

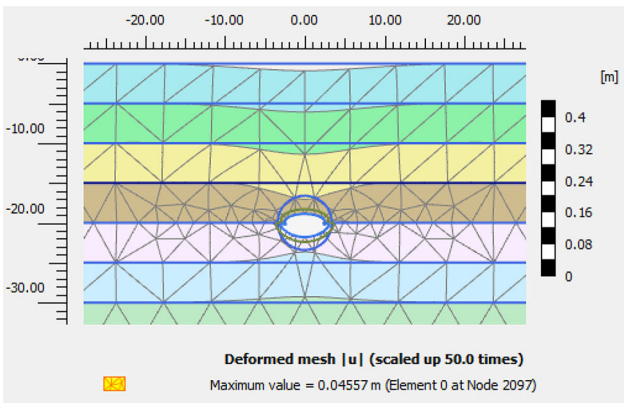


Figure 4. PLAXIS output—deformed mesh after calculations.

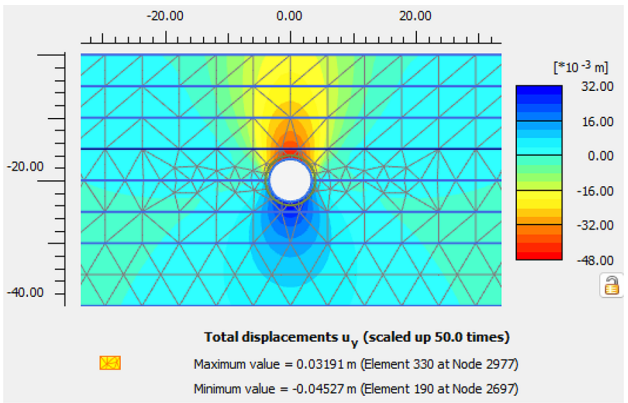


Figure 5. PLAXIS output—total vertical deformations.

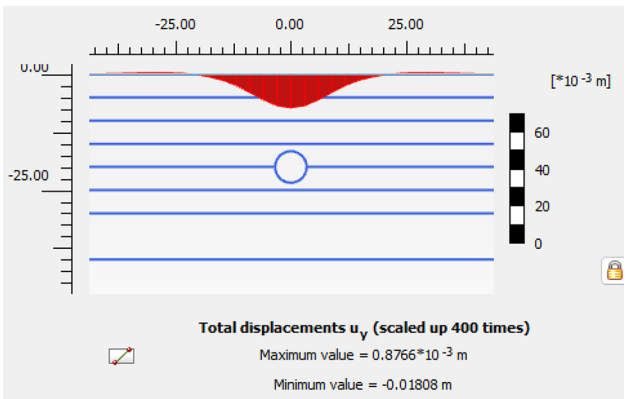


Figure 6. PLAXIS Output- Surface settlement trough.

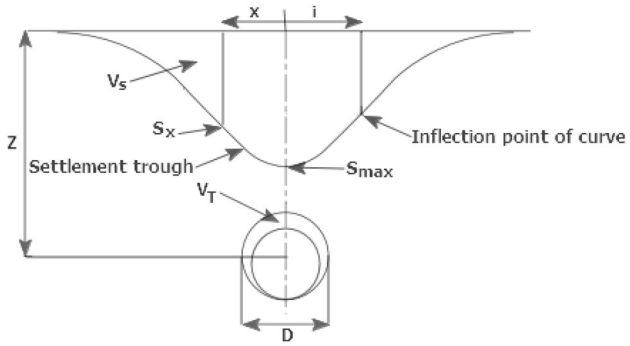


Figure 7. Transverse Surface settlement trough, Gaussian distribution.

from the center line of the tunnel (figure 7) using equation (1) as follows:

$$S_x = S_{max} exp\left(\frac{-x^2}{2i^2}\right) \tag{1}$$

$$S_{max} = 0.313 V_L \frac{D^2}{i} \tag{2}$$

$$V_L = \frac{V_T}{\pi D^2/4} \tag{3}$$

$$i = kz \tag{4}$$

Where, k is the trough width parameter which varies from 0.4 to 0.7 for cohesive soils and 0.2–0.3 for cohesionless soils [37]. z is the depth of tunnel axis. S_{max} is the maximum surface settlement under greenfield condition, V_T is the volume of ground loss around a tunnel cavity per unit length, V_L is volume of ground loss around a tunnel cavity (V_T) with respect to the tunnel volume per unit length and i is the distance between center point of the tunnel and the inflection points. Gaussian curve does not always accurately describe the ground surface settlement in several cases and provide inadequate fit to settlement data above tunnels in sands [4]. Therefore, many researchers modified the Gaussian equation of settlement trough to obtain true surface settlement trough.

Jacobsz *et al* [3] proposed a modified settlement trough equation (5) to accommodate the narrower Gaussian shape settlement trough with k' is the modified trough width parameter and i' is the distance between centre point of the tunnel and the inflection points.

$$S_x = S_{max} exp\left(-\frac{1}{3}\left(\frac{|x|}{i'}\right)^{1.5}\right) \tag{5}$$

$$i' = k'z \tag{6}$$

$$k' = 0.85K - 0.12 \tag{7}$$

Celestino *et al* [38] presented settlement trough as a three-parameter curve equation (8) with S_x, a (length dimension) and b (dimensionless parameter) as three parameters.

$$S_x = \frac{S_{max}}{1 + \left(\frac{|x|}{a}\right)^b} \tag{8}$$

$$i = aB \tag{9}$$

$$B = \left(\frac{b-1}{b+1}\right)^{\frac{1}{b}} \tag{10}$$

$$b = 1 to 3$$

$$\frac{a}{D} = 0.39(z/D) + 0.31 \text{ (For Stiff Clay)} \tag{11}$$

$$\frac{a}{D} = 0.46(z/D) + 0.37 \text{(For Porous Clay)} \tag{12}$$

Vorster *et al* [5] suggested a modified Gaussian curve equation (13) to obtain a better fit to the observed soil settlement, of the following form with a shape function parameter (n) controlling the width of the profile and a parameter α to ensure that i remains the distance to the inflection point.

$$S_x = \frac{nS_{max}}{(n-1) + exp\left[\alpha\left(\frac{x}{i}\right)^2\right]} \tag{13}$$

$$n = e^{\alpha}\frac{2\alpha - 1}{2\alpha + 1} + 1 \tag{14}$$

$$n = 0.5, 1, 1.5$$

Figure 8 compares the surface settlement profiles after excavation of the proposed tunnel from numerical and empirical methods. By idealizing the settlement trough profile obtained by PLAXIS as a Gaussian curve, the settlement trough parameter (i) is found to be 11 m. The corresponding k-value is 0.55. By taking i equals to 11, the Gaussian distribution and modified curves have been

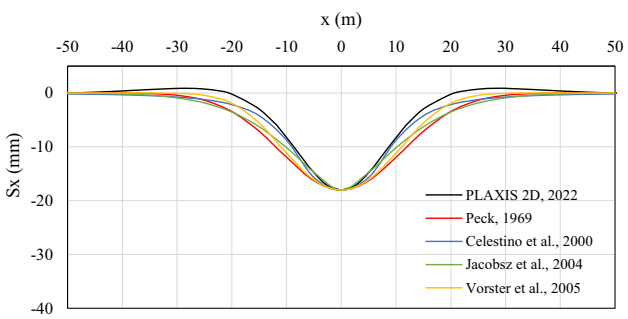


Figure 8. Comparison of surface settlement troughs from PLAXIS 2D and empirical methods.

superimposed on settlement trough obtained by PLAXIS. There is an agreement between results of finite element modelling and empirical methods in terms of shape of profile, but finite element modelling shows a narrow settlement trough, this is due to the layered soil profile under water table [39]. Volume of settlement through is not exactly same as volume loss at tunnel. In this case, volume of settlement through is 1% corresponding to the 1.5% volume loss at tunnel. Mathew and Lehane [13] also observed the same phenomenon and stated that this happened due to effects of the rigidity of the tunnel lining as well as the initial stress state and stiffness of the soil. From figure 8 a minor degree of soil heave is also observed due to dissipation of small negative excess pore pressure in clay layers. Occurrence of heave may also result due to the equal loading and unloading stiffness moduli in the Mohr-Coulomb model [40].

5. Parametric study: results and discussion

Tunnelling-induced ground settlements are influenced by various parameters. Tunnel parameters for which effect on ground settlement analysed are tunnel diameter (D), tunnel axis depth (z) and grout pressure (GP). Soil parameters for which effect on ground settlement analysed are undrained shear strength of soil (s_u) and modulus (E'_{ref}) of soil through which tunnel excavation is done. Influences of each of the above parameters on the tunnelling-induced settlements are analysed. Table 3 highlights the range of parameters that are investigated in the present work. The primary goal of selecting a specific range was to comprehend how these parameters affected the settlement.

5.1 *Effect of tunnel parameters*

Effects of tunnel parameters such as D and z are investigated by maintaining the V_L as 1.5%. Effect of GP has been analysed by maintaining the D as 6.5 m and z as 20 m below the ground surface. Detailed discussion of effects of tunnel parameters is given in the following paragraphs.

5.1a *Tunnel diameter (D) effect* One of the most important factors influencing the ground settlements is the tunnel's size [21]. To study the effect of D on S_x, D is varied from 5.5 m to 8.5 m with an increment of 1 m. Total

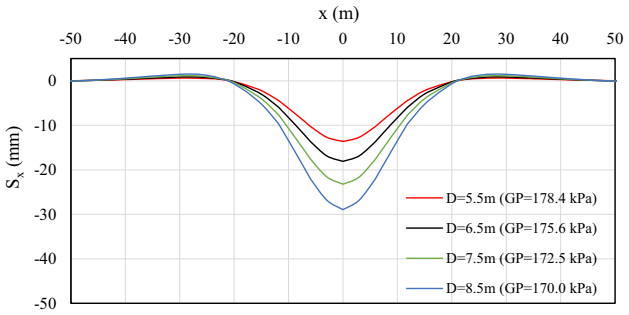


Figure 9. Variation of surface settlement trough with tunnel diameter (D).

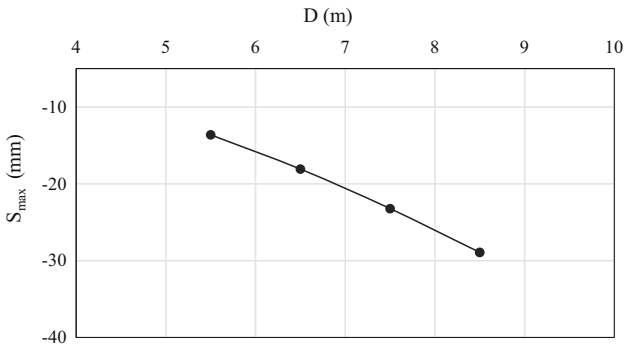


Figure 10. Variation of maximum surface settlement (S_{max}) with tunnel diameter (D)

4 cases have been analysed maintaining the z at 20 m below the ground level. It can be observed from figures 9 and 10, that the increase in D significantly increases the S_x and S_{max} respectively, as reported by many researchers [19, 21]. Maximum surface settlements along with their increment rates for different tunnel diameters are summarized in Table 4. An increase of 112.4% in S_{max} has been found when D increased from 5.5 m to 8.5 m. GP required for maintaining tunnel V_L as 1.5% is also shown in figure 9. This is clear from the GP values that with increase in D less GP is required to maintain the equal V_L. It can also be observed from figure 9, that surface settlement trough peak becomes sharper with increase in the D.

5.1b *Tunnel axis depth (z) effect* The effect of z on S_x has been examined by varying the z from 10 m to 30 m with an increment of 5 m. Total 5 cases have been analysed maintaining the D as 6.5 m. It can be observed from

Table 3. Range of parameters variation for parametric study.

Parameters	Range	Increment	Base Value
Tunnel diameter, D (m)	5.5-8.5	1	6.5
Tunnel depth, z (m)	10-30	5	20
Grout pressure, GP (kPa)	154.2, 175.6, 200 and 237	–	175.6
Undrained shear strength, s_u (kPa)	$0.75s_u$–$1.75s_u$	$0.25s_u$	s_u (90 & 130 for layer 4 & 5 respectively)
Young's modulus, E'_{ref} (kPa)	$0.75E'_{ref}$–$1.75E'_{ref}$	$0.25E'_{ref}$	E'_{ref} ($20*10^3$ & $29*10^3$ for layer 4 & 5 respectively)

Table 4. S_{max} versus D.

Tunnel diameter (D, m)	Maximum surface settlement (S_{max})	Rate of increment (mm/m)
5.5	13.62	–
6.5	18.08	4.46
7.5	23.23	5.15
8.5	28.93	5.7

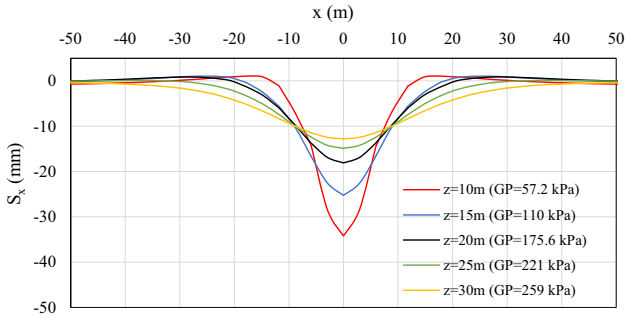


Figure 11. Variation of surface settlement trough with tunnel depth (z).

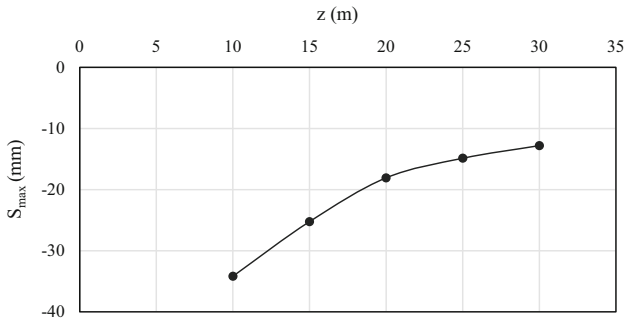


Figure 12. Variation of maximum surface settlement (S_{max}) with tunnel depth (z).

figure 11 that rise in z decreases the S_x near the centreline of the tunnel however away from the centreline opposite phenomenon occurs. Figure 12 shows that increase in z results in significant decrease in S_{max}. Previous researchers also saw a decrease in maximum settlements by increasing the depth of tunnel axis [16, 20, 21, 27].

It can also be observed from figure 11, that width of surface settlement trough increases with increasing z [4]. Table 5 shows the values of maximum surface settlements

Table 5. S_{max} versus z.

Tunnel depth (z, m)	Maximum surface settlement (S_{max})	Rate of decrement (mm/m)
10	34.18	–
15	25.24	1.8
20	18.08	1.4
25	14.87	0.6
30	12.8	0.4

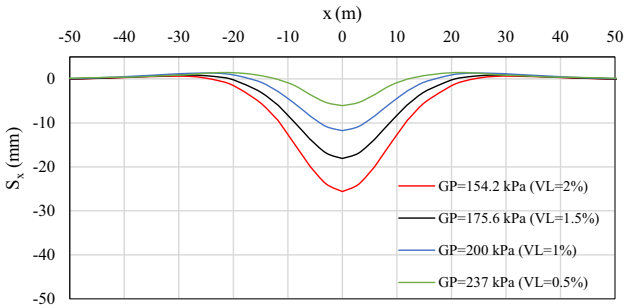


Figure 13. Variation of surface settlement trough with Grout Pressure (GP).

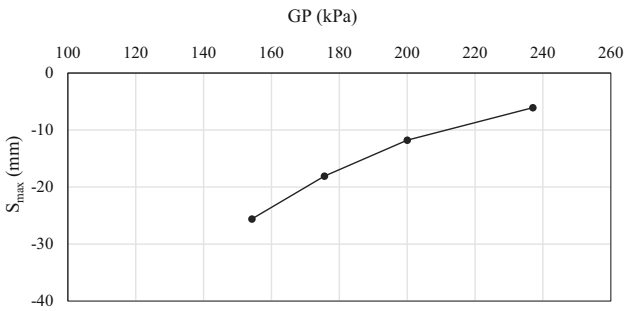


Figure 14. Variation of maximum surface settlement (S_{max}) with Grout Pressure (GP).

Table 6. S_{max} versus GP.

Grout pressure (GP, kPa)	Maximum surface settlement (S_{max})	Rate of decrement (mm/kPa)
154.2	25.6	–
175.6	18.08	0.3514
200	11.77	0.25861
237	6.08	0.15378

and their rate of decrement for each tunnel depth. A decrease of 62.55% in S_{max} has been observed when z increases from 10 m to 30 m. GP required for maintaining V_L 1.5% is also shown in figure 11. This is clear from the GP values that with increase in z more GP is required to maintain the equal V_L. It can also be observed from figure 11, that surface settlement trough peak becomes flatter with increase in the z.

5.1c *Grout pressure (GP) effect* To study the effect of GP on S_x its values are varied as 154.2 kPa, 175.6 kPa, 200 kPa and 237 kPa corresponding to the V_L of 2%, 1.5%, 1% and 0.5% respectively. From the results presented in the figures 13 and 14, it can be observed that the increase in GP results in decrease in V_L which reduces the S_x and S_{max} respectively as foreseen by many researchers [16–18, 21, 41, 42]. Table 6 summarizes the maximum surface settlement values and their rate of decrements for each value of grout pressure. A decrease of 76.25% in S_{max}

has been observed when volume decreases from 2% to 0.5% for the grout pressure values of 154.2 kPa and 237 kPa, respectively.

5.2 *Effect of soil parameters*

Effects of soil parameters were investigated by maintaining the D as 6.5 m and z as 20 m below the ground surface to make it pass through layers 4 and 5. For parametric study parameters of soil from which tunnel pass through have been varied within a practical range. Soil parameters effects are studied in two cases, first by maintaining the V_L as 1.5% and second by maintaining the GP as 175.6 kPa. Detailed discussion of both the cases with respect to soil parameters is given in following paragraphs.

5.2a *Undrained shear strength (s_u) effect* To investigate the effect of s_u its values for layers 4 and 5 have been modified from 0.75 times to 1.75 times with an increment of 0.25 times, with respect to the actual value considered in Table 1. Impact of s_u is illustrated in the figures 15, 16, 17 and 18. It can be observed that increase in s_u results in decrease in S_x and S_{max} in both the cases under constant V_L (1.5%) as well as constant GP (175.6 kPa). Same scenario of decreasing surface settlement with increase in s_u was observed by Kanagaraju and Krishnamurthy [19]. Table 7 summarizes maximum surface settlements along with their

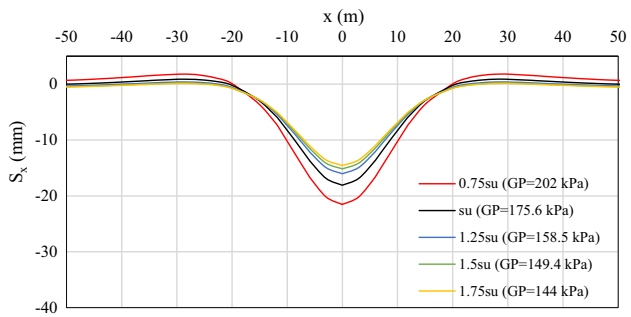


Figure 15. Variation of surface settlement trough with Undrained shear strength (s_u) under constant V_L.

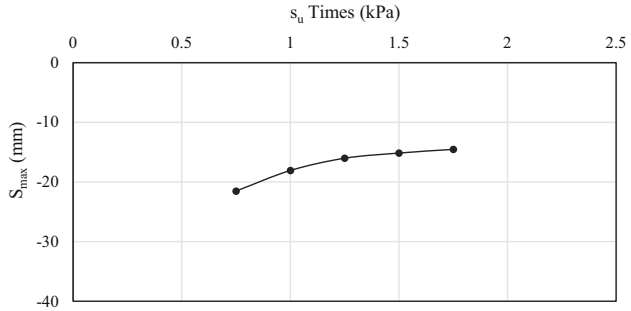


Figure 16. Variation of maximum surface settlement (S_{max}) with Undrained shear strength (s_u) under constant V_L

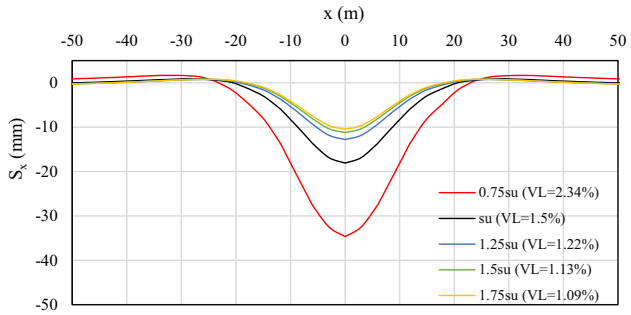


Figure 17. Variation of surface settlement trough with Undrained shear strength (s_u) under constant GP.

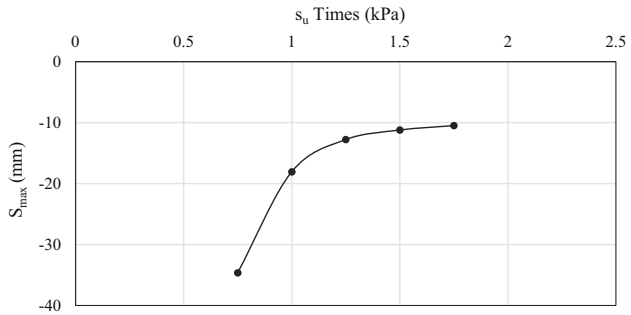


Figure 18. Variation of maximum surface settlement (S_{max}) with Undrained shear strength (s_u) under constant GP.

Table 7. S_{max} versus s_u under constant V_L.

Undrained shear strength (s_u, kPa)	Maximum surface settlement (S_{max})	Rate of decrement (mm/0.25 times s_u)
0.75 s_u	21.53	–
1 s_u	18.08	3.45
1.25 s_u	16.03	2.05
1.5 s_u	15.16	0.87
1.75 s_u	14.55	0.61

rate of decrements for each value of s_u under constant V_L of 1.5%. GP required for maintaining V_L as 1.5% is also shown in figure 15. This is clear from the GP values that with increase in s_u value less GP is required to maintain the equal V_L. Table 8 summarizes maximum surface settlements along with their rate of decrements for each value of

Table 8. S_{max} versus s_u under constant GP.

Undrained shear strength (s_u, kPa)	Maximum surface settlement (S_{max})	Rate of decrement (mm/0.25 times s_u)
0.75 s_u	34.64	–
1 s_u	18.08	16.56
1.25 s_u	12.78	5.3
1.5 s_u	11.22	1.56
1.75 s_u	10.49	0.73

s_u under constant GP of 175.6 kPa. Volume loss caused by constant grout pressure is also shown in figure 17. This is clear from the V_L values that with increase in s_u value V_L decreases with constant GP. When s_u increases upto $1.75s_u$ from $0.75s_u$, a 32.42% and 69.71% decrease in S_{max} have been found under constant V_L and constant GP respectively.

5.2b *Soil modulus (E'_{ref}) effect* To investigate the effect of E'_{ref} its values for layers 4 and 5 have been modified from 0.75 times to 1.75 times with an increment of 0.25 times with respect to the actual value considered in the Table 1. Impact of E'_{ref} is illustrated in figures 19, 20 and 21. It was observed that under constant V_L of 1.5% S_x remains nearly equal for all values of E'_{ref}. GP required for maintaining tunnel V_L 1.5% is also shown in figure 19. This is clear from the GP values that with increase in E'_{ref} value less GP is required to maintain the equal V_L.

Under constant GP of 175.6 kPa settlement decreases with increasing E'_{ref}. Table 9 summarizes the maximum surface settlements and their rate of decrements for each value of E'_{ref} under constant GP. V_L caused by constant GP is also shown in figure 20. This is clear from the V_L that with increase in E'_{ref} V_L decreases under constant GP. Similar trend of decrease in surface settlements and volume loss with increase in modulus was seen by many researchers [19, 21]. A decrease of 32.6% in S_{max} has been observed when E'_{ref} increases from 0.75 times to 1.75 times of its base value.

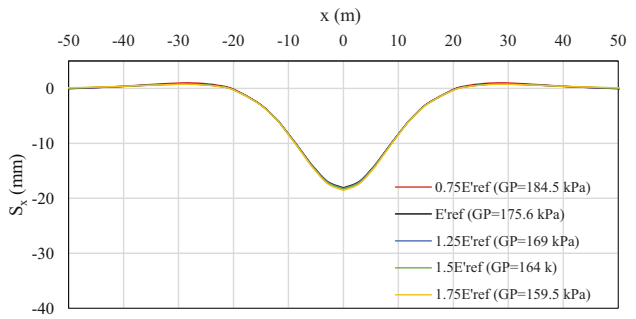


Figure 19. Variation of surface settlement trough with Soil modulus (E'_{ref}) under constant V_L.

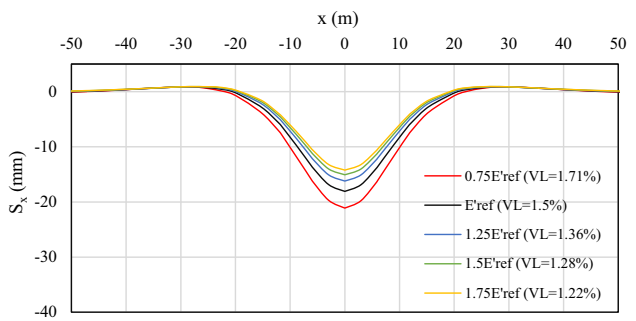


Figure 20. Variation of surface settlement trough with Soil modulus (E'_{ref}) under constant GP.

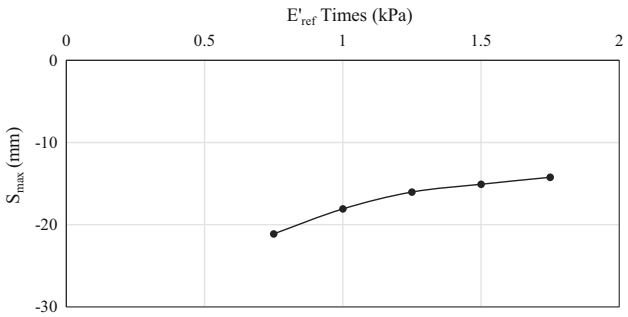


Figure 21. Variation of maximum surface settlement (S_{max}) with Soil modulus (E'_{ref}) under constant GP.

Table 9. S_{max} versus E'_{ref} under constant GP.

Young's modulus (E'_{ref}, kPa)	Maximum surface settlement (S_{max})	Rate of decrement (mm/0.25 times E'_{ref})
0.75 E'_{ref}	21.12	–
1 E'_{ref}	18.08	3.04
1.25 E'_{ref}	16.03	2.05
1.5 E'_{ref}	15.08	0.95
1.75 E'_{ref}	14.23	0.85

6. Conclusion

In the current study TBM tunnelling induced ground settlements under greenfield condition are estimated by simulating step-by-step tunnel construction process in finite element software PLAXIS 2D. A parametric study has been conducted to analyse the effect of different vital tunnel and soil parameters on ground settlements. Prior to conducting the parametric study, the results of numerical model were compared to the existing empirical approaches for validation. Major findings of this study are summarised as follows:

- PLAXIS 2D can provide a reasonable prediction of transverse surface settlements in layered soils of varying stiffnesses.
- Maximum surface settlements and their rate of increments increase with increasing the tunnel diameter under constant volume loss. When the diameter of tunnel was increased from 5.5 m to 8.8 m a significant 112.4% increase in the maximum surface settlement was observed.
- Maximum surface settlements and their rate of decrements decrease with increasing the tunnel depth under constant volume loss. When the tunnel's axial depth was increased from 10 m to 30 m from the ground surface a significant 62.55% decrease in the maximum surface settlement was observed.
- Increase in grout pressure results in decrease in volume loss. Volume loss is an important parameter which affects the maximum surface settlements. Maximum

surface settlements and their rate of decrements decrease with increasing the grout pressure. When the grout pressure was increased from 154.2 kPa to 237 kPa a significant 76.25% decrease in maximum surface settlement was observed.

- Maximum surface settlements and their rate of decrement decrease with increasing the undrained shear strength of soil through which tunnel is excavated under constant volume loss as well as constant grout pressure. When undrained shear strength of soil through which tunnel is excavated is increased from 0.75 to 1.75 times its base value a significant 32.42% and 69.71% decrease in maximum surface settlement was observed under constant volume loss and constant grout pressure, respectively.
- Maximum surface settlements and their rate of decrement decrease with increasing the soil modulus through which tunnel is excavated under constant grout pressure. When soil modulus through which tunnel is excavated is increased from 0.75 to 1.75 times its base value a significant 32.6% decrease in maximum surface settlement was observed under constant grout pressure. However, under constant volume loss, soil modulus does not affect the ground settlements.
- Soil modulus has a little impact on surface settlements as compared to the other parameters.

Above conclusions demonstrate a substantial correlation between surface settlement and tunnel parameters as well as surface settlement and soil parameters, highlighting the significant impact of such modifications on ground stability.

References

[1] Athar M F, Sadique M R, Alsabhan A H and Alam S 2022 Ground Settlement Due to Tunneling in Cohesionless Soil. *Appl. Sci.* 12: 3672

[2] Loganathan N, Poulos H G and Stewart D P 2000 Centrifuge model testing of tunnelling-induced ground and pile deformations. *Geotechnique* 50: 283–294

[3] Jacobsz S W, Standing J R, Mair R J, Hagiwara T and Sugiyama T 2004 Centrifuge modelling of tunnelling near driven piles. *Soils and Foundations* 44: 49–56

[4] Marshall A M, Farrell R, Klar A and Mair R 2012 Tunnels in sands: the effect of size, depth and volume loss on greenfield displacements. *Géotechnique* 62: 385–399

[5] Vorster T E B, Klar A, Soga K and Mair R J 2005 Estimating the Effects of Tunneling on Existing Pipelines. *J. Geotech. Geoenviron. Eng.* 131: 1399–1410

[6] Verruijt A and Booker J R 1996 Surface settlements due to deformation of a tunnel in an elastic half Plane. *Géotechnique* 46: 753–756

[7] Loganathan N and Poulos H G 1998 Analytical prediction for tunnelling induced ground movements in clays. *J. Geotech. Geoenviron. Eng.* 124: 846–856

[8] Fu J, Yang J, Klapperich H and Wang S 2016 Analytical prediction of ground movements due to a nonuniform deforming tunnel. *Int. J. Geomech.* 16: 04015089

[9] Lai H, Zhang J, Zhang L, Chen R and Yang W 2019 A New Method based on Centrifuge Model Test for Evaluating Ground Settlement Induced by Tunnelling. *KSCE Journal of Civil Engineering* 23: 2426–2436

[10] Hong X, Zhang D and Sun Z 2023 Mechanical responses of multi-layered ground due to shallow tunnelling with arbitrary ground surface load. *Frontiers of Structural and Civil Engineering.* 17: 745–762

[11] Sun F, Jin Z, Wang C, Gou C, Li X, Liu C and Yu Z 2023 Case Study on Tunnel Settlement Calculations during Construction Considering Shield Disturbance. *KSCE Journal of Civil Engineering.* 27: 2202–2216

[12] Möller S C and Vermeer P A 2008 On numerical simulation of tunnel installation. *Tunn. Undergr. Space Technol.* 23: 461–475

[13] Mathew G V and Lehane B M 2013 Numerical back-analyses of greenfield settlement during tunnel boring. *Can. Geotech. J.* 50: 145–152

[14] Mathew G V and Lehane B M 2014 Measured and Back Analysed Soil Structure Interaction Effects in a Layered Stratigraphy During Tunnel Boring. *Geotech Geol. Eng.* 32: 873–884

[15] Zheng G, Lu P and Diao Y 2015 Advance speed-based parametric study of greenfield deformation induced by EPBM tunnelling in soft ground. *Computers and Geotechnics* 65: 220–232

[16] Golpasand M R B, Nikudel M R and Uromeihy A 2016 Specifying the real value of volume loss (V_L) and its effect on ground settlement due to excavation of Abuzar tunnel, Tehran. *Bull. Eng. Geol. Environ.* 75: 485–501

[17] Miliziano S and Lillis A D 2019 Predicted and observed settlements induced by the mechanized tunnel excavation of metro line C near S. Giovanni station in Rome. *Tunn. Undergr. Space Technol.* 86: 236–246

[18] Shiau J and Sams M 2019 Relating volume loss and greenfield settlement. *Tunn. Undergr. Space Technol.* 83: 145–152

[19] Kanagaraju R and Krishnamurthy P 2020 Influence of Tunneling in Cohesionless Soil for Different Tunnel Geometry and Volume Loss under Greenfield Condition. *Advances in Civil Engineering* 1–11

[20] Rezaei A H and Adli M A 2020 The Volume Loss: Real Estimation and Its Effect on Surface Settlements Due to Excavation of Tabriz Metro Tunnel. *Geotech. Geol. Eng.* 38: 2663–2684

[21] Aswathy M S, Vinoth M and Mittal A 2022 Impact of Governing Factors on Prediction of Tunneling Induced Surface Settlement in Young Alluvium Deposit. *Indian Geotech. J.* 52: 13–27

[22] Ahmed K S, Sharmin J and Ansary M A 2023 Numerical investigation of tunneling induced surface movement: A case study of MRT Line 1. *Dhaka. Underground Space.* 12: 116–136

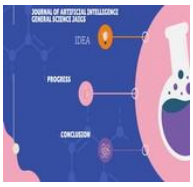
[23] Al-Omari R R, Al-Soud M S and Al-Zuhairi O I 2019 Effect of Tunnel Progress on the Settlement of Existing Piled Foundation. *Studia Geotechnica et Mechanica* 41: 102–113

[24] Islam M S and Iskander M 2021 Twin tunnelling induced ground settlements: A review. *Tunn. Undergr. Space Technol.* 110: 103614
[25] EIMouchi A M, Hassan A M and Amer M I 2018 Performance of an existing raft foundation rested on granular soils due to TBM tunnelling process. *Arabian Journal of Geosciences* 11: 108
[26] Peck R B 1969 Deep excavations and tunneling in soft ground. In: *Proceedings of 7th International Conference on Soil Mechanics and Foundation Engineering, Mexico City, State-of-the-art Volume, pp.* 225–290
[27] Attewell P B and Farmer I W 1974 Ground disturbance caused by shield tunnelling in a stiff, overconsolidated clay. *Engineering Geology* 8: 361–381
[28] Vu M N, Broere W and Bosch J 2016 Volume loss in shallow tunnelling. *Tunn. Undergr. Space Technol.* 59: 77–90
[29] Dias T G S and Bezuijen A 2015 TBM Pressure Models – Observations, Theory and Practice. In: *Proceedings of 15th Pan-American Conference on Soil Mechanics and Geotechnical Engineering - Geotechnical Synergy in Buenos Aires,* pp. 347–374
[30] Wu H N, Shen S L and Yang J 2017 Identification of Tunnel Settlement Caused by Land Subsidence in Soft Deposit of Shanghai. *J. Perform. Constr. Facil.* 31: 04017092
[31] The Metro Railguy Home page https://themetrorailguy.com/delhi-meerut-rrts-information-route-maps-fares-tenders-updates/, Last accessed 03/03/2023
[32] BidAssist Home page https://bidassist.com/delhi-tenders/national-capital-region-transport-corporation/detail-2a42fa4a-5b37-46b5-98b1-7e3bdeb4585b, Last accessed 05/05/2022
[33] Clayton C R 1995 The standard penetration test (SPT): methods and use. *Construction Industry Research and Information Association*. 143
[34] Bowles J E 1988 *Geotechnical and index properties: laboratory testing; settlement and strength correlations, Foundation analysis and design*. 5th edition, New York, McGraw-Hill, 2: 123
[35] PLAXIS 2D- Tutorial Manual. Edition V22
[36] PLAXIS 2D- Reference Manual. Edition V22
[37] O'Reilly M and New B 1982 Settlements above tunnels in the United Kingdom: their magnitude and prediction. In: *Proceedings of Tunnelling '82, Brighton* pp. 173–181
[38] Celestino T B, Gomes R A M P and Bortolucci A A 2000 Errors in Ground Distortions Due to Settlement Trough Adjustment. *Tunn. Undergr. Space Technol.* 15: 97–100
[39] Lee S W 2002 The use of compensation grouting in tunnelling: a case study. *Proceedings of The Institution of Civil Engineers- Geotechnical Engineering* 155: 101–109
[40] Obrzud R F and Truty A 2018 *Introduction The hardening soil model - a practical guidebook. Z Soil.PC 100701 report.* Switzerland: Zaca Service Ltd. pp. 2–10
[41] Fargnoli V, Boldini D and Amorosi A 2013 TBM tunnelling-induced settlements in coarse-grained soils: The case of the new Milan underground line 5. *Tunn. Undergr. Space Technol.* 38: 336–347
[42] Jenck O and Dias D 2003 Numerical analysis of the volume loss influence on building during tunnel excavation. In: *Proceedings of the third international FLAC Symposium-FLAC and FLAC3D Numerical Modeling in Geomechanics, Sudbury, Ontario, Canada* pp. 135–144

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.

ISSN: 3006-4023 (Online), Vol. 2, Issue 1
Journal of Artificial Intelligence General Science (JAIGS)

Journal Homepage: **https://ojs.boulibrary.com/index.php/JAIGS**



Harnessing AI and Gut Microbiome Research for Precision Health

Ritcha Saxena[1]*, Vikas Sharma[2], Ananya Raj Saxena[3], Aakash Patel[4]

[1]*Department of Biomedical Sciences, University of Minnesota School of Medicine, Duluth, Minnesota, USA.

[2] Department of Computer Sciences, Delhi Technological University, New Delhi, India

[3] All Saints University School of Medicine, Dominica

[4] Department of Clinical Sciences, Medical University of the Americas, Nevis, West Indies

Abstract

The gut microbiome's impact on physiological processes, influenced by diet and lifestyle, is profound. Dysbiosis, an imbalance in microbiota composition, is associated with diseases like obesity. This review explores the gut microbiome's role in metabolism and calorie intake, alongside recent AI advancements impacting personalized nutrition. AI has revolutionized microbiome research, especially in multi-omics data analysis. AI-driven approaches enable the integration and interpretation of diverse omics datasets, including genomics, metabolomics, and proteomics, providing comprehensive insights into the gut microbiome's functional dynamics and its impact on host metabolism. These facilitate the identification of microbial biomarkers associated with disease states and dietary interventions, paving the way for personalized nutrition strategies tailored to individual gut microbiome profiles.

AI platforms can also offer tailored dietary recommendations based on microbiome composition and health objectives. Healthcare professionals leverage AI to optimize dietary interventions, promoting gut microbiome modulation and preventing chronic diseases. Challenges like data standardization and privacy persist, yet addressing them is vital for maximizing AI's benefits in health outcomes and precision medicine. Ongoing AI and microbiome research promise to revolutionize personalized nutrition and metabolic health worldwide.

Keywords: Gut microbiome, Calorie intake, Dysbiosis, Short-chain fatty acids, Obesity, Artificial intelligence in healthcare, Multi-omics, Personalized nutrition, Biomarkers

Article Information:
Received: 20/10/2023 *Revised:* 30/11/2024 *Accepted:* 10/01/2024 *Online:* 23/02/2024 *Published*: 23/02/2024

*** Corresponding author:** Ritcha Saxena
Department of Biomedical Sciences, University of Minnesota School of Medicine, 1035 University Dr, Duluth, MN 55812, United States. Email address: rsaxena@d.umn.edu

Background - Understanding the Human Gut Microbiota

The human microbiota is regarded as the assemblage of microorganisms, their genetic material, and their metabolic products that inhabit the human body from birth and are inherited vertically.[1,2] Although all bodily regions are colonized, the gut hosts the most significant microbial populations, a domain extensively researched.[3] The gut microbiome, a complex community of microorganisms residing in the gastrointestinal tract, plays a crucial role in human health. Notably, while survival without a colon is possible, it is impossible without a small intestine, which houses the largest mucosal surfaces in the human body. In the small intestine that houses the majority of gut receptors, nerve and immune cells, food undergoes further digestion and absorption.

The composition of the gut microbiota is dynamic and influenced by various factors, including diet, environment, drug intake, and mode of consumption. The microbiome influences how the host processes and metabolizes dietary nutrients, creating a reciprocal relationship between the gut microbiome and diet. Complex biological and chemical systems interact to modulate an individual's dietary responses, involving the interplay of diet, host, and microbiota.[4,5] The gut microbiome is involved in the metabolism of complex carbohydrates, synthesis of vitamins, and nutrient absorption. It also produces metabolites capable of affecting various physiological mechanisms within the host organism, such as inflammation, metabolism, and immune function. A healthy small intestine harbors a diverse array of microorganisms, predominantly represented by members of the Firmicutes, Bacteroidetes, Actinobacteria, and Proteobacteria phyla, with Firmicutes and Proteobacteria being major phyla among obligate anaerobic bacteria, alongside Bacteroidetes, as noted in multiple metagenomic studies.[6-8] This microbial community reaches its peak in the distal gut, where communities containing up to 100 billion cells per gram can exist for several days, given that colonic transit is considerably longer than that in the small intestine. These microbes are excreted by the body as feces, constituting what is commonly referred to as the gut microbiome, which has been linked to a plethora of ailments and is profoundly influenced by diet and medications.

The gut microbiome is increasingly recognized as a pivotal contributor to regulating host health, encompassing functions such as nutrient provision, defense against pathogens, and fostering immune system maturation.[9] Indeed, the microbiota's role in human health and disease is substantial, sometimes likened to an overlooked organ.[10] Recent studies have broadened our comprehension of various microbiome aspects in humans, shedding light on its influence on diverse physiological processes such as metabolism and energy intake. Clear associations are emerging between the microbiome and its impact on host metabolic processes, with significant implications for human health, particularly in light of the escalating rates of obesity and metabolic syndrome in Western societies.[11] As our understanding of the gut microbiota's physiological role in metabolism advances, it elucidates the mechanisms by which alterations in its composition, both qualitatively and quantitatively, can precipitate metabolic disorders such as obesity.[12-14] This becomes particularly pertinent amidst a global surge in chronic diseases like obesity, which collectively contribute to substantial mortality and morbidity worldwide.[15-17] Indeed, a substantial body of evidence has accumulated, indicating that specific dietary factors and consumption patterns, such as excessive intake of refined carbohydrates coupled with low fiber intake, contribute significantly to the risk of obesity and related conditions, including cardiovascular disease and metabolic syndrome.[18]

Over the past decade, AI and precision medicine have exerted an equal, if not greater, impact on the trajectory of healthcare compared to artificial intelligence. Research in this field has progressed to comprehend how the characterization of health and disease states, along with treatment options for affected individuals, is precisely delineated by integrating multi-omic data with medical history, social/behavioral factors, and environmental insights.[19]

Considering the significant impact of the gut microbiome on metabolism and caloric intake, alongside the advancements in AI applications in medicine, it becomes evident that further exploration of the gut microbiota and the integration of AI could lead to the development of adjunctive treatments beneficially modulating various diseases, particularly those of metabolic nature. While some studies have initiated assessments in these areas, the characterization of the gut microbiota remains incompletely understood, resulting in discordant results in some instances. Here, we offer a comprehensive review of the physiology and composition of the human gut microbiota, with particular emphasis on critically evaluating the current understanding in this field. This review aims to underscore the crucial role of the gut microbiome in regulating calorie intake and metabolism while highlighting how AI is transforming our understanding of the gut microbiome and its impact on these processes. Ultimately, unraveling the complexity of gut microbes and harnessing AI in this domain will pave the way for novel therapies already in development.

Influence of Gut Microbiota on Calorie Intake and Metabolism

Calorie intake, representing the energy derived from food consumption, stands as a fundamental aspect of nutrition and weight management. Historically, it was believed that calorie intake was primarily influenced by the types and quantities of food ingested. Emerging evidence, however, suggests that the composition of the gut microbiome also significantly influences calorie intake and energy balance. Certain bacterial species within the gut microbiome possess metabolic

capabilities facilitating the extraction of calories from the diet. These bacteria can metabolize dietary components typically indigestible by the human host, thus increasing the overall energy yield from food. Also, certain microbial metabolites can influence appetite regulation and satiety signals, thereby affecting overall calorie intake. Alterations in the gut microbiome composition have been associated with increased fat deposition in adipose tissue and exacerbated hepatic inflammatory processes, both contributing to weight gain and metabolic dysfunction. The gut microbiota can also utilize dietary components to produce energy and nutrients for their own growth and proliferation, further influencing the host's energy balance.[20-21]

Studies have demonstrated that the gut microbiome can influence appetite and food preferences through its interaction with the brain-gut axis.[22-23] Certain species of gut bacteria produce signaling molecules that can affect appetite-regulating hormones, such as leptin and ghrelin, thereby influencing feelings of hunger and satiety.[24] The gut microbiome can also modulate taste receptors, influencing our perception of different foods and potentially affecting food choices and calorie intake.[25]

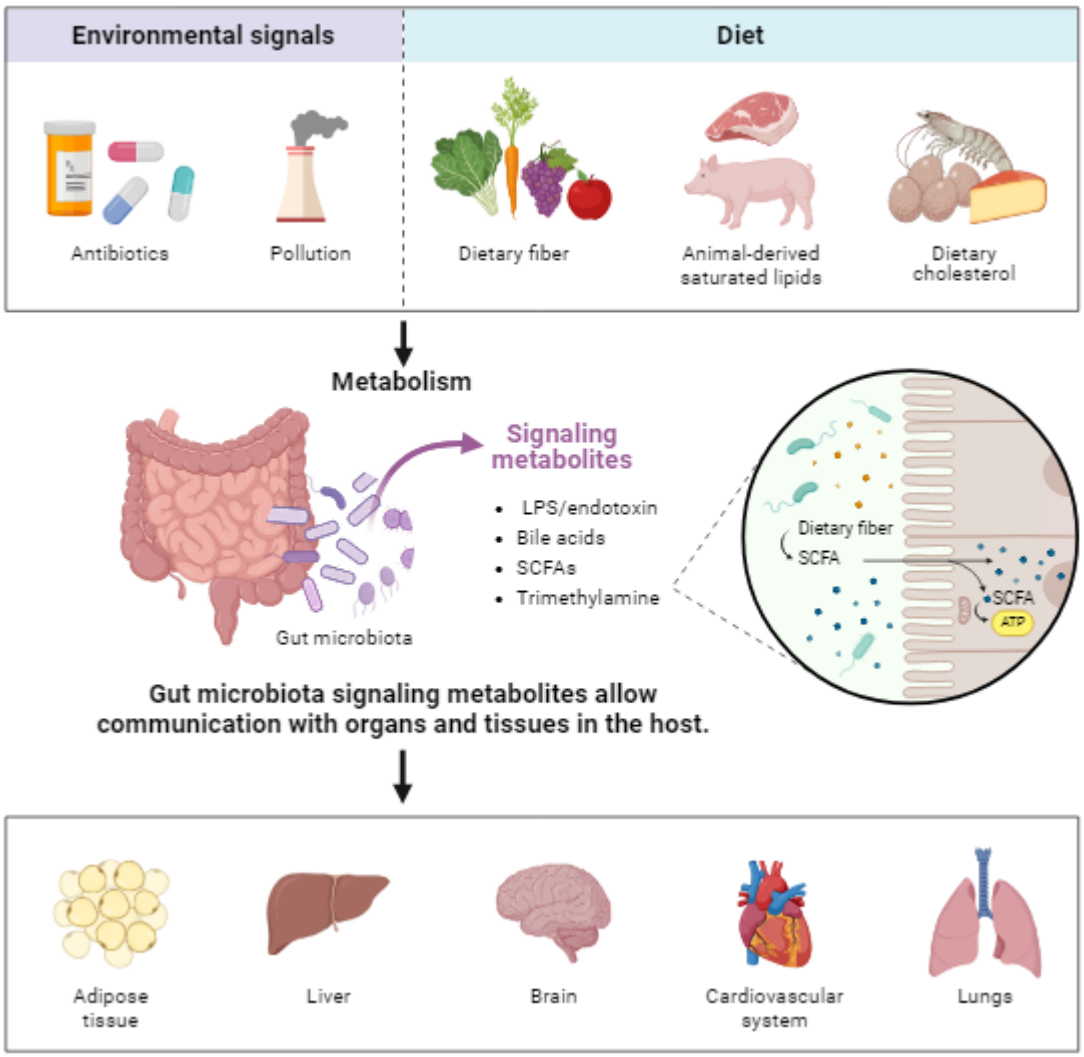


Figure 1: The gut microbiota convert environmental and dietary signals into signaling molecules, affecting host metabolism and inflammatory pathways. Various factors like diet can alter microbiota composition. Microbial metabolites, such as SCFAs and bile acids, signal to various organs, influencing host metabolic pathways and inflammatory responses. The gut microbiota then transforms these inputs into metabolites that signal to different organs and tissues within the host, as illustrated below. Disruptions in microbial balance may worsen the effects of risk factors like the Western diet on obesity and metabolic disorders. Figure adapted from Schroeder BO, Bäckhed F. Nat Med. 2016;22(10):1079-1089. Copyright (2016) by Springer Nature.[37]

The gut microbiota exerts its influence on health outcomes through the metabolization and alteration of various nutrients, giving rise to secondary metabolites with differing retention times, levels of bioactivity, and diverse effects. Examples of these metabolites include short-chain fatty acids (SCFAs), bile acids, hydrogen sulfite, trimethylamine and

lipopolysachharide endotoxins. Numerous metabolites produced by the gut microbiota have the ability to alter host metabolism and immune system pathways.[26-28] They can either promote health, such as SCFAs, or exacerbate disease, such as bile acids or hydrogen sulfite.[28,30] Further, the gut microbiota ferments complex sugars and dietary fibers that evade digestion by human enzymes.[31] Phyla such as Firmicutes and Bacteroidetes, which exhibit some tolerance to oxygen, encode a variety of enzymes necessary for breaking down complex carbohydrates, including fiber and resistant starch, which would otherwise remain indigestible.[32] This fermentation process yields SCFAs, including acetate, propionate, and butyrate, serving as an energy source for the host and influencing energy metabolism, glucose homeostasis and inflammation of tissues such as adipose.[33] In particular, SCFAs have been demonstrated to influence calorie consumption and storage by regulating hunger, raising energy expenditure, and promoting fat oxidation.[34,35] Several receptors, including free fatty acid receptor 3 (FFAR3 or GPR41) and niacin receptor 1 (GPR109A), have been identified for SCFAs, highlighting their diverse physiological roles in inflammation, atherosclerosis, and pancreatic β-cell growth and insulin secretion.[36]

Finally, the composition of the gut microbiome exhibits significant variability among patients, influenced by various factors including genetics, diet, lifestyle, and environmental exposures. Significant interindividual variations exist in the diversity, composition, and function of the microbiome.[38,39] Many observational and intervention studies have shown how host genetic, epigenetic, and lifestyle factors including nutrition shape and influence variance in the gut microbiota.[40-43]

Dysbiosis and Disease: The Impact of Gut Microbiota Composition on Metabolic Disorders

The composition, function, and diversity of the gut microbiome are closely associated with an individual's predisposition to a wide range of diseases, many of which are linked to diet. Dysbiosis, an imbalance in the gut microbiota, has been implicated in various diseases affecting different parts of the body.[44,45] It is now understood that alterations in the gut microbiome and microbial functions play a significant role in the pathogenesis of various diseases, including cancer and metabolic disorders. Such changes in the gut microbiota composition and function, when combined with traditional genetic predispositions and lifestyle factors, can contribute to the development and progression of these conditions.[46,47] On the other hand, reduced calorie consumption and better metabolic health are associated with a varied and balanced gut microbiome that is rich in beneficial bacteria. Studies have indicated that individuals with healthier metabolic profiles tend to have a more diverse microbiome composition, characterized by a greater abundance of beneficial bacterial species.[48] Studies also show that in the gut microbiomes of obese individuals, there tends to be a higher proportion of genes encoding membrane transport functions and those involved in butyrate synthesis. Conversely, genes linked to cofactor, vitamin, and nucleotide metabolism or transcription are more often depleted.[49,50] A study of human and mouse microbiota correlated obesity with differences in the relative abundance of two dominant bacterial divisions and showed that obese individuals have an increased capacity to harvest energy from the diet.[51] Relative abundance of certain bacterial phyla, particularly Bacteroidetes and Firmicutes, is altered in the gut microbiomes of obese individuals. Specifically, Bacteroidetes are often less common, while Firmicutes are more prevalent in obese individuals compared to lean individuals.[52] The development of gut microbe-targeted therapies by reversing dysbiosis of the microbiota, inhibiting microbial enzymes or genetically engineered probiotics, has been suggested to be feasible and efficacious.[53,54]

Comprehending the intricate connection between the intestinal microbiota and energy intake holds substantial ramifications for dietary patterns and well-being. Approaches directed towards fostering a robust intestinal microbiota, like ingesting a varied array of plant-derived foods abundant in fiber, probiotics, and prebiotics, might aid in bolstering optimal metabolic performance and managing energy consumption. a balanced diet that includes a sufficient amount of fiber promotes microbial fermentation of fiber into SCFAs, fosters microbial variability and stands as one mechanism by which high fiber consumption prevents weight gain, potentially outweighing inherited predispositions to obesity.[55-57] Hence, it is plausible that a broader dietary regimen supplemented with appropriate nourishment could rectify the intraluminal milieu in instances of dysbiosis.[58-60] Moreover, delving deeper into the mechanisms governing the influence of the gut microbiome on appetite control and metabolic activity could pave the way for tailored interventions targeting obesity and other metabolic conditions.

Therefore, the gut microbiome assumes a pivotal role in regulating energy balance and calorie intake by influencing appetite, metabolism, and food preferences. Encouraging a varied and well-balanced gut microbiome through dietary adjustments and lifestyle modifications could introduce innovative strategies for weight management and enhancing metabolic well-being. The effective fusion of dietary elements with the gut microbiome also holds considerable potential to transform disease management paradigms, offering personalized dietary suggestions, lifestyle alterations, or even the optimization of the gut microbiota tailored to individual needs and health objectives.[61,62]

Integration of AI into Bioinformatics

The rapid growth of computer hardware and software in the healthcare industry in recent years has facilitated the digitization of health data, opening up new avenues for the development of computational models and opportunities to utilize AI systems for deriving insights from data.[63]

In a recent paper, the authors underscored the "unprecedented opportunities" for boosting the expertise of professionals and the supportive function AI serves in alleviating human constraints like weariness and lack of focus, while also avoiding the dangers of mechanical fallibility. More significantly, the article stresses the importance of judicious implementation of these technological advancements while recognizing their significant potential.[64] The digitization of health-related data and the swift adoption of technology are driving transformation and advancement in the utilization of AI in healthcare.[65-67] Access to healthcare, environment, diet, and way of life all have a big impact on our personal health. The advent of wearable technology and other medical devices has enabled the monitoring and quantification of these behavioral, social, and external aspects. These variables collectively contribute to approximately 60% of our health determinants, encompassing physiological, psychological, behavioral, and socioeconomic data. In contrast, genetics constitute only 10% of these determinants, with medical history accounting for the remaining 30%.[68]

Integration of AI into bioinformatics has enabled the analysis of large datasets, a task often challenging with conventional statistical methodologies. AI is applied across various domains including medical diagnosis, support for therapeutic techniques, prediction of disease onset risks, and other related areas.[69] AI is even transforming our understanding of the gut microbiota and how it affects metabolism and calorie intake. A systems-wide mechanistic perspective that takes into account pre-disease, disease, and health states is necessary to gain a thorough knowledge of the dynamics within networks associated with nutrition and health benefits. This need emphasizes the need for cutting-edge techniques and procedures that can both quantify the effects of dietary changes in healthy persons and facilitate comparisons with patients who are diseased.[70] AI algorithms are finding increasing applications in personalized nutrition research, offering valuable insights and support for dietary activities, disease risk assessment related to food and nutrient patterns, and exploration of supplementation research.

AI Applications in Personalized Nutrition and Gut Microbiome Analysis

The primary aim of precision medicine is to tailor care to the specific needs of each patient.[71,72] The momentum behind precision nutrition is rapidly growing, underscoring its increasing significance as we come to acknowledge its breadth. de Toro-Martín and colleagues elaborate on precision nutrition methodologies, which extend far beyond genetics to encompass elements such as dietary patterns, eating habits, physical activity, the microbiome, and the metabolome.[73] Personalized nutrition stands on the cusp of a significant transformation with the advent of AI, which, through its capacity to analyze vast datasets, discern patterns, and make predictions, is opening up new avenues for personalized nutrition and precision medicine. Personalized nutrition is witnessing a surge, with advancements in AI and gut microbiome analysis revolutionizing our approach to diet and health. These personalized nutrition approaches have the potential to spearhead the development of information-processing representations of digestion, absorption, and metabolism.[74-76]

By harnessing advanced computational techniques, AI empowers researchers to uncover intricate patterns and relationships within large datasets that were previously difficult to discern. The escalating utilization of AI algorithms in this domain reflects scientific advancement and is increasingly becoming not just an asset, but a necessity in the quest for valuable outcomes. Through sophisticated AI systems, healthcare professionals can leverage data analysis capabilities to discern patterns and offer personalized diagnoses and treatment plans for various gastrointestinal conditions.

Numerous studies have highlighted that the underlying variability in the gut microbiome can be attributed to various factors, including diet, genetics, gender, age, lifestyle, environmental exposure, epigenetics, drugs, and geography. These intricacies may complicate the associations between the microbiota and human diseases.[77-80] Given these complexities, there is a pressing need to develop advanced computational methods capable of efficiently extracting key information from vast, heterogeneous, and complex multi-omics data. AI algorithms can leverage data from gut microbiome analysis, genetic makeup, and lifestyle to offer tailored dietary recommendations. This data-driven approach takes into consideration the interplay between the gut microbiome, diet, and environmental factors, ensuring individuals receive personalized nutrition guidance tailored to their specific requirements.[81,82]

In gut microbiome research, this is a primary application of AI involves analyzing microbiome composition and diversity. AI in biomedical sciences offers valuable tools and techniques for gathering, structuring, and scrutinizing vast biological datasets like nutritional, genomic, and related data.[83] AI systems are capable of analyzing high-throughput sequencing data to classify and determine the number of different microbial species in the gut. More research is now underway, including longitudinally individualized multi-omics profiling—which includes genomes, metabolomics, and proteomics—thanks to developments in high-throughput technology.[84] This method improves comprehension of the relationships between the gut microbiota and the host and provides a thorough grasp of human metabolism.[85-87] AI diagnostics employ advanced algorithms to scrutinize diverse data sources, including stool samples, blood analyses, and health records. This enables researchers to profile the microbiome of individuals and populations with unparalleled precision and detail.

Furthermore, AI algorithms can analyze the interactions between gut microbes and host physiology to elucidate the mechanisms underlying their effects on calorie intake and metabolism. The application of AI in biomedical nutrition research addresses the need for effective examination and understanding of the complex interactions between nutrition and human physiology, particularly in the context of the gut microbiome.[88-90] Important microbial metabolites, signaling pathways, and host-microbiome interactions that affect energy expenditure, calorie utilization, and appetite regulation can be identified using machine learning models. Utilizing a comprehensive approach facilitates the discovery of important markers that contribute to the development and progression of diseases such as colon cancer or irritable bowel syndrome (IBS). Timely detection and precise diagnoses are pivotal in ameliorating patient outcomes and enhancing treatment efficacy.

Predictive modeling driven by AI is also used to create individualized food plans based on each person's unique gut microbiota composition. The individuality of each person's gut microbiome is a critical facet of gut health. By examining the gut microbiome in conjunction with AI-powered diagnostics, researchers and healthcare professionals can better understand and address individual needs AI systems analyze a wide range of factors, including gut microbiota, lifestyle, and environmental factors, to provide deep insights and create customized treatment plans. These recommendations consider various reactions to specific nutrients obtained from food, resulting from the complex interplay between nutrients and biological processes.[91] These includes the correlation between external influences like food habits and physical activity, and internal factors like microbiota-metabolome interactions and genetics.[92] For instance, AI systems are able to provide personalized food recommendations based on an individual's genetic composition by combining genetic data with nutrition databases.[93] The variability in response to therapies and dietary advice among individuals is reflected in biomarker levels.[94] These anticipatory findings can be utilized to comprehend the complex regulatory mechanisms of dietary interventions at the convergence of immunity, metabolism, and gut flora. AI algorithms can forecast the potential impacts of various dietary strategies on an individual's gut microbiota and metabolic health by combining microbiome details with data on eating habits, lifestyle factors, and metabolic markers. This personalized approach enables precise dietary suggestions aimed at optimizing gut microbiota composition and regulating calorie intake.

AI can also be employed to explore dietary preferences and trends in relation to the gut microbiota's composition. AI systems can uncover correlations between specific food components and shifts in the gut microbiome's structure by analyzing extensive databases of microbiome profiles and dietary intake records. This information can guide the development of dietary advice and interventions aimed at nurturing a balanced gut microbiome and managing calorie consumption. In this direction, AI-driven personalized nutrition is poised for ongoing enhancements, potentially facilitating the creation of a global network capable of actively overseeing and enhancing each person's nutrient intake.[95]

AI's integration into gut microbiome research represents a groundbreaking frontier for advancing the understanding of the intricate interplays among the microbiome, calorie consumption, and metabolism. Through sophisticated computational analysis and data-driven methodologies, AI empowers scientists to navigate the complex mechanisms within the microbiome ecosystem. AI further enables the design of targeted interventions tailored to optimize metabolic well-being, paving the way for personalized nutrition.

Potential of AI in Advancing Gut Microbiome Research

In recent years, there have been several notable advancements in the field of AI and gut microbiome research, particularly concerning calorie intake and metabolism.

The field of microbiome-based therapeutics is rapidly expanding. A growing body of evidence suggests that modifying the gut microbiota through different therapeutic regimens such as dietary adjustments, fecal microbiota transplantation, or supplementation with probiotics and prebiotics holds promise for preventing or treating various illnesses.[96,97] AI is playing a crucial role in advancing the development of microbiome-based treatments for metabolic diseases like obesity. By identifying key microbial taxa and metabolites associated with metabolic health, researchers can tailor therapies such as microbial transplantation to optimize gut microbiome function and improve metabolic outcomes. Understanding the mechanisms underlying the functions of gut microbiota functions also opens up possibilities for developing highly effective probiotics with targeted benefits.[90,98]

AI algorithms focused on microbiome-host interactions are unraveling the complex connections between the gut microbiome and host physiology, shedding light on hunger regulation, energy metabolism, and gut-brain communication. AI has played a pivotal role in advancing our mechanistic understanding of the gut microbiome by identifying novel microbial genomes and proteins from uncultured species, leveraging vast amounts of gut metagenomic data. Additionally, machine learning techniques can predict protein structures based on unidentified protein sequences, aiding in enzyme design and drug development.[99] These breakthroughs enhance our understanding of how the gut microbiota influences calorie intake and metabolism, paving the way for innovative therapeutic strategies.

Using machine learning to find biomarkers connected to certain dietary treatments or health outcomes is greatly facilitated by advancements in bioinformatics and artificial intelligence. Researchers are increasingly leveraging machine learning techniques to uncover microbial biomarkers associated with metabolic health and disease. In earlier investigations of the microbiota, the interpretability of the models allowed for the simultaneous identification of biomarkers linked to disease, providing valuable biological insights into the data. This methodology enables the discovery of new insights into illnesses and facilitates the identification of viable treatment approaches.[100] For instance, Yu et al. (2017) employed the minimum redundancy–maximum relevance (mRMR) feature selection method to identify an optimal set of 20 microbial genes predictive of colorectal cancer, with four validated genes distinguishing patients from controls across different ethnic cohorts. These findings underscore the potential of machine learning in identifying microbial biomarkers for the early diagnosis of colorectal cancer.[101] Additionally, machine learning has been widely applied to the identification of microbial biomarkers for the purpose of evaluating the risk of disease and creating tailored treatments for gut microorganisms. Deep learning has shown great promise in a number of disciplines, such as AlphaGo and AlphaFold, which are applied to analyze large-scale microbiome datasets, offering new insights into microbial communities and their interactions with the host.[99,102] AI systems have the capability to identify microbial signatures indicative of an individual's risk for conditions like obesity, insulin resistance, and other metabolic disruptions by analyzing vast microbiome datasets. This facilitates early detection and the development of personalized treatment strategies.

Integrative methodologies in omics are essential for discerning the connections between the gut microbiota and human diseases. Single-omics analysis often provides only a partial view of the intricate biological systems. Recent research in gut microbiome highlights the vast potential of AI in integrating and analyzing diverse data sets.[103,104] AI algorithms enable the integration of various omics data types, including genomics, metabolomics, and microbiomics, leading to a deeper understanding of the gut microbiome and its role in metabolism. Despite challenges such as high dimensionality and limited sample sizes, deep learning has proven effective in handling and integrating multi-omics data.[105,106] By offering a comprehensive understanding of the molecular processes underlying both metabolic health and illness, these integrated approaches simplifying the identification of novel treatment objectives and strategies

The latest advancements in artificial intelligence and gut microbiome research are significantly enhancing our understanding of the complex interplay between the microbiota, caloric intake, and metabolism. Through the utilization of sophisticated computational techniques and data analytics platforms, there is the potential to revolutionize the development of health platforms, customize forthcoming dietary guidelines to optimize health outcomes, and expedite the integration of these guidelines into clinical practice. By leveraging AI-driven methodologies, researchers can glean novel insights into metabolic health and devise state-of-the-art strategies for personalized dietary and therapeutic interventions.

Challenges, opportunities, and practical implications of using AI in personalized nutrition

Several challenges must be overcome to effectively utilize AI in exploring the connection between the gut microbiota and calorie intake and to implement it in precision medicine for personalized therapy. Despite progress, this field still faces obstacles that hinder seamless integration into real-world healthcare. One major challenge is the establishment of gut microbiota biobanks using AI technology for scientific research, which looms as a significant hurdle on the horizon.[107] The incorporation of AI in healthcare introduces several challenges, including those related to multimodal data integration, federated learning, model performance, bias, and privacy concerns. Multimodal data integration entails merging information from diverse sources such as medical records, imaging studies, and genetic data, which can be intricate and demand advanced computational methods. Federated learning, which involves training machine learning models across multiple decentralized data sources, necessitates significant advancements in privacy-preserving techniques, large-scale machine learning, and distributed optimization to ensure data privacy and security while maintaining effective model performance.[108]

Ensuring consistent data quality and standardization across diverse sources is imperative. Discrepancies in data collection, sequencing methodologies, and preprocessing can introduce biases that affect the accuracy and reliability of AI predictions. Evaluating the credibility and reliability of test outcomes generated by AI techniques is a critical challenge.[109] The domain context and applications of AI in healthcare are closely intertwined with considerations of fairness and protected qualities. However, challenges arise due to limitations in the representativeness of training samples derived from patient data. Often, the sample sizes are insufficient to fully capture the diverse variations among patients and the complexity of their health conditions. The challenges are compounded by the curse of dimensionality, especially when dealing with gut microbiome data, where high dimensionality and limited sample sizes can hinder traditional AI methods.[110] This challenge is termed as bias, which remains one of the significant hurdles in the application of AI in healthcare. Data bias occurs when the training data used to develop AI models is not representative of the population it aims to serve. This can lead to inaccuracies or unfair outcomes, particularly for underrepresented groups, and can exacerbate existing health disparities.[111] Further research in the field of medical AI is essential to understand and address fairness and bias in AI models constructed using historical patient data. Collaboration between the biomedical and AI communities is crucial to tackle this challenge effectively. To mitigate bias and overfitting in AI models, it is imperative to validate algorithms derived from retrospective and single-center studies

and promote validation across multiple centers and in prospective studies. Standardizing data processing methods would not only streamline clinical data collection but also facilitate predictive analysis. This approach enhances the robustness and applicability of AI models, ensuring their effectiveness across diverse settings.[112] Standardizing data also promotes seamless communication among medical facilities, physicians, academic institutions, and data scientists involved in data evaluation. This facilitates quick and consistent identification of missing data for individual patients. Furthermore, ensuring data uniformity across healthcare organizations ensures the quality of patient care and facilitates monitoring of their records, thereby enabling evidence-based recommendations to be formulated.

The lack of clarity in complex AI algorithms, like deep learning, presents challenges in grasping underlying mechanisms. The demand for individuals adept at bridging computational and clinical fields has surged with the advent of new data-driven technologies. Ensuring model clarity is pivotal for effectively translating research findings into actionable insights. This highlights the necessity to train individuals with broad expertise in computer science, statistics, biology, nutrition, biomedicine, mathematics, and data science. Integrating concepts from statistics, machine learning, bioinformatics, mathematics, and computer science into training programs for students, trainees, and professionals in data science is imperative. Workshops and training events designed to unite researchers from diverse backgrounds can bolster response predictability.[113] The training of the new professionals should reflect the cutting-edge knowledge guided by the change in day to day informatics challenges.[114] Trainers stand to gain valuable insights from trainings and tutorials, obtaining up-to-date knowledge on the latest technology, data standards, and utilization procedures for infrastructure services.[115]

Robust AI models that accurately capture the dynamics of the gut microbiome across populations and dietary habits require large, diverse datasets for training. However, obtaining such databases can be costly and logistically challenging. Exploring alternative methods to facilitate the development of high-performing AI systems, such as generating synthetic data with realistic variations or leveraging simulated environments, is imperative. While early research shows promise, further investigation into AI is warranted.[116] Furthermore, despite AI's ability to identify relationships, distinguishing causation from correlation remains a challenge. Mere concatenation of raw data or model outputs from each perspective overlooks opportunities to explore potential connections and interactions among diverse multi-omic data. To establish causality and validate AI-driven predictions, longitudinal research and experimental interventions are essential.

Concerns regarding data ownership, consent, and privacy arise when AI utilizes gut microbiome data for personalized recommendations. The issue of trust in AI-based technologies, especially among the senior population, has not been fully addressed. Recent research has showcased the potential of new computational algorithms to identify individuals within public or commercial health databases.[117] A crucial consideration is that health data should not be used to discriminate against individuals in determining insurance costs or employment status.[95] The use of genetic data poses extra challenges since the drivers of prediction algorithms remain uncertain.[117] Unauthorized access to and misuse of electronic medical records present additional concerns. Improper access to medical records can lead to privacy breaches.[118] Establishing a secure and tightly regulated environment for data storage, management, and exchange is imperative. This necessitates the adoption of new technologies, collaborative efforts, and the development of new business models and legislation. Transparency, informed consent, and data security are essential for maintaining trust and protecting individuals' rights. One approach that AI system engineers can explore to address ongoing privacy concerns is the use of generative data. Generative models enable the creation of synthetic patient data that is realistic yet unrelated to actual individuals.[119,120] This makes machine learning possible without requiring the long-term use of actual patient data.

The cost of AI technology may render it inaccessible to certain individuals and healthcare facilities. Residents in rural or isolated areas, lacking adequate internet connectivity or technological infrastructure, may face limited access.[121] Thus, in order to improve access to social, economic, and educational opportunities, digital equity must be promoted. It is imperative to surmount financial obstacles and augment accessibility to guarantee universal availability of customized dietary plans and metabolic health treatments.

Collaboration across disciplines, rigorous study designs, transparent reporting, and adherence to ethical standards are effective strategies for addressing these challenges. Overcoming these challenges would allow AI to fulfill its potential in revolutionizing our understanding of how the gut microbiome regulates energy intake and metabolism. This enhanced understanding could lead to more effective approaches for improving metabolic health and preventing chronic diseases.

Conclusion

Recent studies have shed light on the multifaceted roles of the gut microbiota in human health, emphasizing microbial diversity as a principal indicator of overall well-being. However, substantial inter-individual variability exists, influenced by environmental factors and dietary quality.[122-126] Nutrition and diet are recognized as significant determinants of an individual's health, with a growing emphasis on understanding how dietary patterns and specific diets impact an individual’s health. This shift in research focus—from analyzing the effects of individual nutrients to examining the functional aspects

of dietary patterns—reflects a deeper understanding of the complex interactions between diet, gut microbiota, and human health.[127-129]

Due to the significant role that diet plays in influencing gut microbiota composition and host metabolism, our understanding of actions patients can take to enhance their systemic health is evolving.[130,131] The relationships observed in microbial diversity emphasize the significance of consuming a nourishing, diverse diet. They support the idea that dietary supplements containing particular nutrients should not be viewed as replacements for a balanced and healthy diet.[132,133] Nutrition and dietary habits exert a more substantial influence on the prognosis of chronic diseases such as diabetes, obesity, IBS and colorectal cancer compared to genetic and environmental factors. These elements impact metabolism and inflammation.[134-137]

Despite advancements, unanswered questions remain regarding the precise composition of an optimal health-promoting microbiome and strategies for cultivating such flora in individuals with initially diverse microbiotas. Standardized dietary approaches may not always yield desired outcomes. Discrepancies in the effectiveness of lifestyle interventions may be attributed to variations in gut flora among individuals.[77,137,138] It is noted that patients exhibit highly personalized reactions of their microbiomes to diverse meals, depending on their previous dietary experiences.[139] Consequently, owing to advancements in AI and gut microbiome research, personalized nutrition and healthcare are on the brink of a transformative shift. The use of AI in health research and initiatives has surged in recent years, facilitating the integration of various clinical data elements, establishing connections between biorepositories and clinical data, and bridging the gap between clinical data and pharmaceutical research and development.[140-143] AI-driven platforms have the potential to enhance metabolic health outcomes and prevent chronic diseases such as obesity by providing personalized food recommendations based on an individual's gut microbiome composition and metabolic profile. Future investigations should focus on elucidating the intricate functions of the gut microbiota in these processes. These findings may lead to the utilization of the gut microbiota and its metabolites in the treatment of metabolic disorders in the future. Probiotics and prebiotics exemplify customized interventions offered by microbiome-based therapies driven by AI, targeting specific microbial imbalances to provide more effective and personalized treatment options.

Moreover, AI algorithms can detect early signs of metabolic dysfunction through the analysis of microbial biomarkers. This enables timely intervention and the prevention of chronic disease. The advancement of medications designed to modulate the gut microbiota and the development of personalized treatment plans are facilitated by our expanding knowledge of how the gut microbiome influences drug metabolism. Early detection of metabolic abnormalities facilitated by AI-driven diagnostic technologies is another avenue through which patient outcomes can be enhanced and healthcare resources optimized.

Furthermore, in the realm of medical education, the incorporation of AI into personalized nutrition holds significant promise. For instance, the integration of digital pathology and AI, already in practice in several medical education settings, greatly enhances students' comprehension of pathophysiology concepts.[144] This integration can further offers insights into the complex relationship between cellular morphology and physiological processes, including those influenced by the gut microbiota. Through the use of AI-driven diagnostic tools, aspiring physicians can delve into the subtle connections between tissue pathology and systemic health, acquiring valuable insights to inform personalized nutrition strategies aimed at optimizing metabolic well-being for both themselves and their future patients.

Future research endeavors should aim to elucidate the roles of nonbacterial microorganisms in the human gut and the shifts in bacterial composition associated with different disease states. Future developments in AI and gut microbiome research have the potential to revolutionize precision medicine, public health initiatives, and healthcare efficiency. These advancements may offer unprecedented opportunities to enhance metabolic health and overall well-being on a global scale. Ultimately, leveraging AI across the field of gut microbiota research will facilitate advancements in precision nutrition, personalized medicine, and the development of next-generation probiotics tailored to individual patients.[77,27,145] While AI and precision medicine hold immense promise, efforts to test, validate, and refine treatment protocols require continued dedication. Overcoming challenges such as obtaining sufficient high-quality labeled data for algorithm training, navigating regulatory and privacy considerations, and adopting standardized data formats are crucial steps for researchers to address. This review highlights several key points for further consideration and investigation. Promising strides are being made in surmounting the obstacles faced by AI in healthcare, paving the way for continued growth and advancement in AI's role in both personal and population health.

References

1. Arumugam M, Raes J, Pelletier E, et al. Enterotypes of the human gut microbiome [published correction appears in Nature. 2011 Jun 30;474(7353):666] [published correction appears in Nature. 2014 Feb 27;506(7489):516]. Nature. 2011;473(7346):174-180. doi:10.1038/nature09944

2. Korpela K, de Vos WM. Early life colonization of the human gut: microbes matter everywhere. Curr Opin Microbiol. 2018;44:70-78. doi:10.1016/j.mib.2018.06.003
3. Lloyd-Price J, Mahurkar A, Rahnavard G, et al. Strains, functions and dynamics in the expanded Human Microbiome Project [published correction appears in Nature. 2017 Oct 12;:]. Nature. 2017;550(7674):61-66. doi:10.1038/nature23889)
4. Kolodziejczyk AA, Zheng D, Elinav E. Diet-microbiota interactions and personalized nutrition. Nat Rev Microbiol. 2019;17(12):742-753. doi:10.1038/s41579-019-0256-8
5. Hughes RL, Kable ME, Marco M, Keim NL. The Role of the Gut Microbiome in Predicting Response to Diet and the Development of Precision Nutrition Models. Part II: Results. Adv Nutr. 2019;10(6):979-998. doi:10.1093/advances/nmz049
6. Seekatz AM, Schnizlein MK, Koenigsknecht MJ, et al. Spatial and Temporal Analysis of the Stomach and Small-Intestinal Microbiota in Fasted Healthy Humans. mSphere. 2019;4(2):e00126-19. Published 2019 Mar 13. doi:10.1128/mSphere.00126-19
7. Leite GGS, Weitsman S, Parodi G, et al. Mapping the Segmental Microbiomes in the Human Small Bowel in Comparison with Stool: A REIMAGINE Study. Dig Dis Sci. 2020;65(9):2595-2604. doi:10.1007/s10620-020-06173-x
8. Kaur K, Khatri I, Akhtar A, Subramanian S, Ramya TNC. Metagenomics analysis reveals features unique to Indian distal gut microbiota [published correction appears in PLoS One. 2020 Dec 1;15(12):e0243397]. PLoS One. 2020;15(4):e0231197. Published 2020 Apr 8. doi:10.1371/journal.pone.0231197
9. Gill SR, Pop M, Deboy RT, et al. Metagenomic analysis of the human distal gut microbiome. Science. 2006;312(5778):1355-1359. doi:10.1126/science.1124234
10. O'Hara AM, Shanahan F. The gut flora as a forgotten organ. EMBO Rep. 2006;7(7):688-693. doi:10.1038/sj.embor.7400731
11. Harakeh SM, Khan I, Kumosani T, et al. Gut Microbiota: A Contributing Factor to Obesity. Front Cell Infect Microbiol. 2016;6:95. Published 2016 Aug 30. doi:10.3389/fcimb.2016.00095
12. Jumpertz R, Le DS, Turnbaugh PJ, et al. Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. Am J Clin Nutr. 2011;94(1):58-65. doi:10.3945/ajcn.110.010132
13. Rowland I, Gibson G, Heinken A, et al. Gut microbiota functions: metabolism of nutrients and other food components. Eur J Nutr. 2018;57(1):1-24. doi:10.1007/s00394-017-1445-8
14. Gomes AC, Hoffmann C, Mota JF. The human gut microbiota: Metabolism and perspective in obesity. Gut Microbes. 2018;9(4):308-325. doi:10.1080/19490976.2018.1465157
15. Bauer UE, Briss PA, Goodman RA, Bowman BA. Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. Lancet. 2014;384(9937):45-52. doi:10.1016/S0140-6736(14)60648-6
16. van Dijk SJ, Tellam RL, Morrison JL, Muhlhausler BS, Molloy PL. Recent developments on the role of epigenetics in obesity and metabolic disease. Clin Epigenetics. 2015;7:66. Published 2015 Jul 11. doi:10.1186/s13148-015-0101-5
17. Heitmann BL, Westerterp KR, Loos RJ, et al. Obesity: lessons from evolution and the environment. Obes Rev. 2012;13(10):910-922. doi:10.1111/j.1467-789X.2012.01007.x
18. Liu S. Intake of refined carbohydrates and whole grain foods in relation to risk of type 2 diabetes mellitus and coronary heart disease. J Am Coll Nutr. 2002;21(4):298-306. doi:10.1080/07315724.2002.10719227
19. Ziegelstein RC. Personomics and Precision Medicine. Trans Am Clin Climatol Assoc. 2017;128:160-168.
20. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. Science. 2005;307(5717):1915-1920. doi:10.1126/science.1104816
21. Neyrinck AM, Etxeberria U, Taminiau B, et al. Rhubarb extract prevents hepatic inflammation induced by acute alcohol intake, an effect related to the modulation of the gut microbiota. Mol Nutr Food Res. 2017;61(1):10.1002/mnfr.201500899. doi:10.1002/mnfr.201500899
22. Barrea L, Muscogiuri G, Annunziata G, et al. From gut microbiota dysfunction to obesity: could short-chain fatty acids stop this dangerous course?. Hormones (Athens). 2019;18(3):245-250. doi:10.1007/s42000-019-00100-0
23. van de Wouw M, Schellekens H, Dinan TG, Cryan JF. Microbiota-Gut-Brain Axis: Modulator of Host Metabolism and Appetite. J Nutr. 2017;147(5):727-745. doi:10.3945/jn.116.240481
24. Fetissov SO. Role of the gut microbiota in host appetite control: bacterial growth to animal feeding behaviour. Nat Rev Endocrinol. 2017;13(1):11-25. doi:10.1038/nrendo.2016.150
25. Turner A, Veysey M, Keely S, Scarlett C, Lucock M, Beckett EL. Interactions between Bitter Taste, Diet and Dysbiosis: Consequences for Appetite and Obesity. Nutrients. 2018;10(10):1336. Published 2018 Sep 20. doi:10.3390/nu10101336
26. Nicholson JK, Holmes E, Kinross J, et al. Host-gut microbiota metabolic interactions. Science. 2012;336(6086):1262-1267. doi:10.1126/science.1223813
27. Rooks MG, Garrett WS. Gut microbiota, metabolites and host immunity. Nat Rev Immunol. 2016;16(6):341-352. doi:10.1038/nri.2016.42

28. WJ, Hase K. Gut microbiota-generated metabolites in animal health and disease. Nat Chem Biol. 2014;10(6):416-424. doi:10.1038/nchembio.1535
29. Hughes RL, Marco ML, Hughes JP, Keim NL, Kable ME. The Role of the Gut Microbiome in Predicting Response to Diet and the Development of Precision Nutrition Models-Part I: Overview of Current Methods. Adv Nutr. 2019;10(6):953-978. doi:10.1093/advances/nmz022
30. Ni Y, Li J, Panagiotou G. A Molecular-Level Landscape of Diet-Gut Microbiome Interactions: Toward Dietary Interventions Targeting Bacterial Genes. mBio. 2015;6(6):e01263-15. Published 2015 Oct 27. doi:10.1128/mBio.01263-15
31. Lamichhane S, Sen P, Dickens AM, Orešič M, Bertram HC. Gut metabolome meets microbiome: A methodological perspective to understand the relationship between host and microbe. Methods. 2018;149:3-12. doi:10.1016/j.ymeth.2018.04.029
32. Sims IM, Ryan JL, Kim SH. In vitro fermentation of prebiotic oligosaccharides by Bifidobacterium lactis HN019 and Lactobacillus spp. Anaerobe. 2014;25:11-17. doi:10.1016/j.anaerobe.2013.11.001.
33. Janssen AW, Kersten S. Potential mediators linking gut bacteria to metabolic health: a critical view. J Physiol. 2017;595(2):477-487. doi:10.1113/JP272476
34. Murugesan S, Nirmalkar K, Hoyo-Vadillo C, García-Espitia M, Ramírez-Sánchez D, García-Mena J. Gut microbiome production of short-chain fatty acids and obesity in children. Eur J Clin Microbiol Infect Dis. 2018;37(4):621-625. doi:10.1007/s10096-017-3143-0
35. Liu JL, Segovia I, Yuan XL, Gao ZH. Controversial Roles of Gut Microbiota-Derived Short-Chain Fatty Acids (SCFAs) on Pancreatic β-Cell Growth and Insulin Secretion. Int J Mol Sci. 2020;21(3):910. Published 2020 Jan 30. doi:10.3390/ijms21030910
36. Ohira H, Tsutsui W, Fujioka Y. Are Short Chain Fatty Acids in Gut Microbiota Defensive Players for Inflammation and Atherosclerosis?. J Atheroscler Thromb. 2017;24(7):660-672. doi:10.5551/jat.RV17006
37. Schroeder BO, Bäckhed F. Signals from the gut microbiota to distant organs in physiology and disease. Nat Med. 2016;22(10):1079-1089. doi:10.1038/nm.4185
38. Koppel N, Balskus EP. Exploring and Understanding the Biochemical Diversity of the Human Microbiota. Cell Chem Biol. 2016;23(1):18-30. doi:10.1016/j.chembiol.2015.12.008
39. Yatsunenko T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. Nature. 2012;486(7402):222-227. Published 2012 May 9. doi:10.1038/nature11053
40. Bonder MJ, Kurilshikov A, Tigchelaar EF, et al. The effect of host genetics on the gut microbiome. Nat Genet. 2016;48(11):1407-1412. doi:10.1038/ng.3663
41. Benson AK, Kelly SA, Legge R, et al. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. Proc Natl Acad Sci U S A. 2010;107(44):18933-18938. doi:10.1073/pnas.1007028107
42. David LA, Maurice CF, Carmody RN, et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature. 2014;505(7484):559-563. doi:10.1038/nature12820
43. De Filippo C, Cavalieri D, Di Paola M, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc Natl Acad Sci U S A. 2010;107(33):14691-14696. doi:10.1073/pnas.1005963107
44. Musso G, Gambino R, Cassader M. Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. Annu Rev Med. 2011;62:361-380. doi:10.1146/annurev-med-012510-175505
45. Zhang YJ, Li S, Gan RY, Zhou T, Xu DP, Li HB. Impacts of gut bacteria on human health and diseases. Int J Mol Sci. 2015;16(4):7493-7519. Published 2015 Apr 2. doi:10.3390/ijms16047493
46. Hope ME, Hold GL, Kain R, El-Omar EM. Sporadic colorectal cancer--role of the commensal microbiota. FEMS Microbiol Lett. 2005;244(1):1-7. doi:10.1016/j.femsle.2005.01.029
47. Vijay-Kumar M, Aitken JD, Carvalho FA, et al. Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. Science. 2010;328(5975):228-231. doi:10.1126/science.1179721
48. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci Transl Med. 2009;1(6):6ra14. doi:10.1126/scitranslmed.3000322
49. Greenblum S, Turnbaugh PJ, Borenstein E. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. Proc Natl Acad Sci U S A. 2012;109(2):594-599. doi:10.1073/pnas.1116053109
50. Ferrer M, Ruiz A, Lanza F, et al. Microbiota from the distal guts of lean and obese adolescents exhibit partial functional redundancy besides clear differences in community structure. Environ Microbiol. 2013;15(1):211-226. doi:10.1111/j.1462-2920.2012.02845.x
51. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444(7122):1027-1031. doi:10.1038/nature05414
52. Turnbaugh PJ, Hamady M, Yatsunenko T, et al. A core gut microbiome in obese and lean twins. Nature. 2009;457(7228):480-484. doi:10.1038/nature07540

53. Yuvaraj S, Al-Lahham SH, Somasundaram R, Figaroa PA, Peppelenbosch MP, Bos NA. E. coli-produced BMP-2 as a chemopreventive strategy for colon cancer: a proof-of-concept study. Gastroenterol Res Pract. 2012;2012:895462. doi:10.1155/2012/895462
54. Chen K, Zhu Y, Zhang Y, et al. A probiotic yeast-based immunotherapy against Clostridioides difficile infection. Sci Transl Med. 2020;12(567):eaax4905. doi:10.1126/scitranslmed.aax4905
55. Chambers ES, Preston T, Frost G, Morrison DJ. Role of Gut Microbiota-Generated Short-Chain Fatty Acids in Metabolic and Cardiovascular Health. Curr Nutr Rep. 2018;7(4):198-206. doi:10.1007/s13668-018-0248-8
56. Fernandes J, Su W, Rahat-Rozenbloom S, Wolever TM, Comelli EM. Adiposity, gut microbiota and faecal short chain fatty acids are linked in adult humans. Nutr Diabetes. 2014;4(6):e121. Published 2014 Jun 30. doi:10.1038/nutd.2014.23
57. Menni C, Jackson MA, Pallister T, Steves CJ, Spector TD, Valdes AM. Gut microbiome diversity and high-fibre intake are related to lower long-term weight gain. Int J Obes (Lond). 2017;41(7):1099-1105. doi:10.1038/ijo.2017.66
58. Llewellyn SR, Britton GJ, Contijoch EJ, et al. Interactions Between Diet and the Intestinal Microbiota Alter Intestinal Permeability and Colitis Severity in Mice. Gastroenterology. 2018;154(4):1037-1046.e2. doi:10.1053/j.gastro.2017.11.030
59. Chiba M, Abe T, Tsuda H, et al. Lifestyle-related disease in Crohn's disease: relapse prevention by a semi-vegetarian diet. World J Gastroenterol. 2010;16(20):2484-2495. doi:10.3748/wjg.v16.i20.2484
60. Lewis JD, Abreu MT. Diet as a Trigger or Therapy for Inflammatory Bowel Diseases. Gastroenterology. 2017;152(2):398-414.e6. doi:10.1053/j.gastro.2016.10.019
61. Bashiardes S, Godneva A, Elinav E, Segal E. Towards utilization of the human genome and microbiome for personalized nutrition. Curr Opin Biotechnol. 2018;51:57-63. doi:10.1016/j.copbio.2017.11.013
62. Eetemadi A, Rai N, Pereira BMP, Kim M, Schmitz H, Tagkopoulos I. The Computational Diet: A Review of Computational Methods Across Diet, Microbiome, and Health. Front Microbiol. 2020;11:393. Published 2020 Apr 3. doi:10.3389/fmicb.2020.00393
63. Hinton G. Deep Learning-A Technology With the Potential to Transform Health Care. JAMA. 2018;320(11):1101-1102. doi:10.1001/jama.2018.11100
64. Matheny ME, Whicher D, Thadaney Israni S. Artificial Intelligence in Health Care: A Report From the National Academy of Medicine. JAMA. 2020;323(6):509-510. doi:10.1001/jama.2019.21579
65. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7
66. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial Intelligence in Surgery: Promises and Perils. Ann Surg. 2018;268(1):70-76. doi:10.1097/SLA.0000000000002693
67. Mayo RC, Leung J. Artificial intelligence and deep learning - Radiology's next frontier?. Clin Imaging. 2018;49:87-88. doi:10.1016/j.clinimag.2017.11.007
68. McGinnis JM, Williams-Russo P, Knickman JR. The case for more active policy attention to health promotion. Health Aff (Millwood). 2002;21(2):78-93. doi:10.1377/hlthaff.21.2.78
69. Kaput J. Lessons from application of data science strategies in nutritional research. Am J Clin Nutr. 2020;111(1):4-5. doi:10.1093/ajcn/nqz284
70. Verma M, Hontecillas R, Abedi V, et al. Modeling-Enabled Systems Nutritional Immunology. Front Nutr. 2016;3:5. Published 2016 Feb 16. doi:10.3389/fnut.2016.00005
71. Aronson SJ, Rehm HL. Building the foundation for genomics in precision medicine. Nature. 2015;526(7573):336-342. doi:10.1038/nature15816
72. Anderson JL, Horne BD, Stevens SM, et al. Randomized trial of genotype-guided versus standard warfarin dosing in patients initiating oral anticoagulation. Circulation. 2007;116(22):2563-2570. doi:10.1161/CIRCULATIONAHA.107.737312
73. de Toro-Martín J, Arsenault BJ, Després JP, Vohl MC. Precision Nutrition: A Review of Personalized Nutritional Approaches for the Prevention and Management of Metabolic Syndrome. Nutrients. 2017;9(8):913. Published 2017 Aug 22. doi:10.3390/nu90809130
74. Abrahams M, Matusheski NV. Personalised nutrition technologies: a new paradigm for dietetic practice and training in a digital transformation era. J Hum Nutr Diet. 2020;33(3):295-298. doi:10.1111/jhn.12746
75. Barrow M, Bell L, Bell C. Transforming personalized nutrition practice. Nutr Rev. 2020;78(12):1046-1051. doi:10.1093/nutrit/nuaa012
76. Adams SH, Anthony JC, Carvajal R, et al. Perspective: Guiding Principles for the Implementation of Personalized Nutrition Approaches That Benefit Health and Function. Adv Nutr. 2020;11(1):25-34. doi:10.1093/advances/nmz086
77. Sharma M, Li Y, Stoll ML, Tollefsbol TO. The Epigenetic Connection Between the Gut Microbiome in Obesity and Diabetes. Front Genet. 2020;10:1329. Published 2020 Jan 15. doi:10.3389/fgene.2019.01329
78. Deschasaux M, Bouter KE, Prodan A, et al. Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. Nat Med. 2018;24(10):1526-1531. doi:10.1038/s41591-018-0160-1
79. Wu H, Esteve E, Tremaroli V, et al. Metformin alters the gut microbiome of individuals with treatment-naive type 2 diabetes, contributing to the therapeutic effects of the drug. Nat Med. 2017;23(7):850-858. doi:10.1038/nm.4345

80. He Y, Wu W, Zheng HM, et al. Regional variation limits applications of healthy gut microbiome reference ranges and disease models [published correction appears in Nat Med. 2018 Sep 24;:]. Nat Med. 2018;24(10):1532-1535. doi:10.1038/s41591-018-0164-x
81. Boland M, Alam F, Bronlund J. Modern Technologies for Personalized Nutrition. 2019; doi:10.1016/B978-0-12-816403-7.00006-4.
82. Forster H, Walsh MC, Gibney MJ, Brennan L, Gibney ER. Personalised nutrition: the role of new dietary assessment methods. Proc Nutr Soc. 2016;75(1):96-105. doi:10.1017/S0029665115002086
83. Lecroq T, Soualmia LF. Managing large-scale genomic datasets and translation into clinical practice. Yearb Med Inform. 2014;9(1):212-214. Published 2014 Aug 15. doi:10.15265/IY-2014-0039
84. Saxena R, Saxena RR, Saxena AR. Microbiomics in the Molecular Era: A Bird's Eye View into the Future of Personalized Medicine. Acta Scientific Microbiology. 2018;1(8):34-39.
85. Lloyd-Price J, Arze C, Ananthakrishnan AN, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature. 2019;569(7758):655-662. doi:10.1038/s41586-019-1237-9
86. Zhou W, Sailani MR, Contrepois K, et al. Longitudinal multi-omics of host-microbe dynamics in prediabetes. Nature. 2019;569(7758):663-671. doi:10.1038/s41586-019-1236-x
87. Integrative HMP (iHMP) Research Network Consortium. The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. Cell Host Microbe. 2014;16(3):276-289. doi:10.1016/j.chom.2014.08.014.
88. Devika NT, Raman K. Deciphering the metabolic capabilities of Bifidobacteria using genome-scale metabolic models. Sci Rep. 2019;9(1):18222. Published 2019 Dec 3. doi:10.1038/s41598-019-54696-9
89. Shima H, Masuda S, Date Y, et al. Exploring the Impact of Food on the Gut Ecosystem Based on the Combination of Machine Learning and Network Visualization. Nutrients. 2017;9(12):1307. Published 2017 Dec 1. doi:10.3390/nu9121307
90. Mohammed A, Guda C. Application of a hierarchical enzyme classification method reveals the role of gut microbiome in human metabolism. BMC Genomics. 2015;16 Suppl 7(Suppl 7):S16. doi:10.1186/1471-2164-16-S7-S16
91. Celis-Morales C, Lara J, Mathers JC. Personalising nutritional guidance for more effective behaviour change. Proc Nutr Soc. 2015;74(2):130-138. doi:10.1017/S0029665114001633
92. Qi L. Personalized nutrition and obesity. Ann Med. 2014;46(5):247-252. doi:10.3109/07853890.2014.891802
93. Lee A, Mavaddat N, Wilcox AN, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors [published correction appears in Genet Med. 2019 Feb 21;:]. Genet Med. 2019;21(8):1708-1718. doi:10.1038/s41436-018-0406-9
94. van Ommen B, Keijer J, Kleemann R, et al. The challenges for molecular nutrition research 2: quantification of the nutritional phenotype. Genes Nutr. 2008;3(2):51-59. doi:10.1007/s12263-008-0084-3
95. Verma M, Hontecillas R, Tubau-Juni N, Abedi V, Bassaganya-Riera J. Challenges in Personalized Nutrition and Health. Front Nutr. 2018;5:117. Published 2018 Nov 29. doi:10.3389/fnut.2018.00117.
96. Jia W, Li H, Zhao L, Nicholson JK. Gut microbiota: a potential new territory for drug targeting. Nat Rev Drug Discov. 2008;7(2):123-129. doi:10.1038/nrd2505
97. Nagpal R, Mainali R, Ahmadi S, et al. Gut microbiome and aging: Physiological and mechanistic insights. Nutr Healthy Aging. 2018;4(4):267-285. Published 2018 Jun 15. doi:10.3233/NHA-170030
98. O'Toole PW, Marchesi JR, Hill C. Next-generation probiotics: the spectrum from probiotics to live biotherapeutics. Nat Microbiol. 2017;2:17057. Published 2017 Apr 25. doi:10.1038/nmicrobiol.2017.57
99. Senior AW, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning. Nature. 2020;577(7792):706-710. doi:10.1038/s41586-019-1923-7
100. Wilmanski T, Rappaport N, Earls JC, et al. Blood metabolome predicts gut microbiome α-diversity in humans. Nat Biotechnol. 2019;37(10):1217-1228. doi:10.1038/s41587-019-0233-9
101. Yu J, Feng Q, Wong SH, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut. 2017;66(1):70-78. doi:10.1136/gutjnl-2015-309800309800
102. Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search. Nature. 2016;529(7587):484-489. doi:10.1038/nature16961
103. Zeevi D, Korem T, Zmora N, et al. Personalized Nutrition by Prediction of Glycemic Responses. Cell. 2015;163(5):1079-1094. doi:10.1016/j.cell.2015.11.001
104. Singh A, Shannon CP, Gautier B, et al. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics. 2019;35(17):3055-3062. doi:10.1093/bioinformatics/bty1054
105. Tsubaki M, Tomii K, Sese J. Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. Bioinformatics. 2019;35(2):309-318. doi:10.1093/bioinformatics/bty535
106. Zhang L, Lv C, Jin Y, et al. Deep Learning-Based Multi-Omics Data Integration Reveals Two Prognostic Subtypes in High-Risk Neuroblastoma. Front Genet. 2018;9:477. Published 2018 Oct 18. doi:10.3389/fgene.2018.00477
107. Zhou T, Wang M, Ma H, Li X, Heianza Y, Qi L. Dietary Fiber, Genetic Variations of Gut Microbiota-derived Short-chain Fatty Acids, and Bone Health in UK Biobank. J Clin Endocrinol Metab. 2021;106(1):201-210.

doi:10.1210/clinem/dgaa740
108. Wang F, Preininger A. AI in Health: State of the Art, Challenges, and Future Directions. Yearb Med Inform. 2019;28(1):16-26. doi:10.1055/s-0039-1677908
109. Lu Y, Stathopoulou T, Vasiloglou MF, et al. An Artificial Intelligence-Based System for Nutrient Intake Assessment of Hospitalised Patients. Annu Int Conf IEEE Eng Med Biol Soc. 2019;2019:5696-5699. doi:10.1109/EMBC.2019.8856889
110. Clarke R, Ressom HW, Wang A, et al. The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. Nat Rev Cancer. 2008;8(1):37-49. doi:10.1038/nrc2294
111. Parikh RB, Teeple S, Navathe AS. Addressing Bias in Artificial Intelligence in Health Care. JAMA. 2019;322(24):2377-2378. doi:10.1001/jama.2019.18058
112. Carter SM, Rogers W, Win KT, Frazer H, Richards B, Houssami N. The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. Breast. 2020;49:25-32. doi:10.1016/j.breast.2019.10.001
113. Vodovotz Y, Xia A, Read EL, et al. Solving Immunology?. Trends Immunol. 2017;38(2):116-127. doi:10.1016/j.it.2016.11.006
114. Kannry J, Sengstack P, Thyvalikakath TP, et al. The Chief Clinical Informatics Officer (CCIO): AMIA Task Force Report on CCIO Knowledge, Education, and Skillset Requirements. Appl Clin Inform. 2016;7(1):143-176. Published 2016 Mar 16. doi:10.4338/ACI-2015-12-R-0174
115. Snoek HM, Eijssen LMT, Geurts M, Vors C, Brown KA, Zimmermann MJ, van't Veer PP. Advancing food, nutrition, and health research in Europe by connecting and building research infrastructures in a DISH-RI: Results of the EuroDISH project. Trends Food Sci Tech. 2018;73:58–66. doi:10.1016/j.tifs.2017.12.015
116. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA. 2017;318(22):2199-2210. doi:10.1001/jama.2017.14585
117. Na L, Yang C, Lo CC, Zhao F, Fukuoka Y, Aswani A. Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets From Which Protected Health Information Has Been Removed With Use of Machine Learning. JAMA Netw Open. 2018;1(8):e186040. Published 2018 Dec 7. doi:10.1001/jamanetworkopen.2018.6040
118. Mossotto E, Ashton JJ, O'Gorman L, et al. GenePy - a score for estimating gene pathogenicity in individuals using next-generation sequencing data. BMC Bioinformatics. 2019;20(1):254. Published 2019 May 16. doi:10.1186/s12859-019-2877-3
119. Yoon J, Drumright LN, van der Schaar M. Anonymization Through Data Synthesis Using Generative Adversarial Networks (ADS-GAN). IEEE J Biomed Health Inform. 2020;24(8):2378-2388. doi:10.1109/JBHI.2020.2980262
120. Baowaly MK, Lin CC, Liu CL, Chen KT. Synthesizing electronic health records using improved generative adversarial networks. J Am Med Inform Assoc. 2019;26(3):228-241. doi:10.1093/jamia/ocy142
121. Barker H. Global economic inequality and health. Med Confl Surviv. 2020;36(4):368-374. doi:10.1080/13623699.2020.1848577
122. Saxena R, Yakyomovych O. Gut Microbiome, A Link between Nutrition, Physiology, and Pathology: Insights into Current Status and Future Directions. Int J Collab Res Intern Med Public Health. 2023,15(3):1-5
123. Claesson MJ, Jeffery IB, Conde S, et al. Gut microbiota composition correlates with diet and health in the elderly. Nature. 2012;488(7410):178-184. doi:10.1038/nature11319
124. Pérez Martínez G, Bäuerl C, Collado MC. Understanding gut microbiota in elderly's health will enable intervention through probiotics. Benef Microbes. 2014;5(3):235-246. doi:10.3920/BM2013.0079.
125. Davenport ER, Mizrahi-Man O, Michelini K, Barreiro LB, Ober C, Gilad Y. Seasonal variation in human gut microbiome composition. PLoS One. 2014;9(3):e90731. Published 2014 Mar 11. doi:10.1371/journal.pone.0090731
126. O'Toole PW, Jeffery IB. Gut microbiota and aging. Science. 2015;350(6265):1214-1215. doi:10.1126/science.aac8469
127. Hills RD Jr, Pontefract BA, Mishcon HR, Black CA, Sutton SC, Theberge CR. Gut Microbiome: Profound Implications for Diet and Disease. Nutrients. 2019;11(7):1613. Published 2019 Jul 16. doi:10.3390/nu11071613
128. Jin Q, Black A, Kales SN, Vattem D, Ruiz-Canela M, Sotos-Prieto M. Metabolomics and Microbiomes as Potential Tools to Evaluate the Effects of the Mediterranean Diet. Nutrients. 2019;11(1):207. Published 2019 Jan 21. doi:10.3390/nu11010207
129. Lara KM, Levitan EB, Gutierrez OM, et al. Dietary Patterns and Incident Heart Failure in U.S. Adults Without Known Coronary Disease. J Am Coll Cardiol. 2019;73(16):2036-2045. doi:10.1016/j.jacc.2019.01.067
130. Gentile CL, Weir TL. The gut microbiota at the intersection of diet and human health. Science. 2018;362(6416):776-780. doi:10.1126/science.aau5812
131. Zmora N, Suez J, Elinav E. You are what you eat: diet, health and the gut microbiota. Nat Rev Gastroenterol Hepatol. 2019;16(1):35-56. doi:10.1038/s41575-018-0061-2
132. Zhernakova A, Kurilshikov A, Bonder MJ, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science. 2016;352(6285):565-569. doi:10.1126/science.aad3369
133. Chen F, Du M, Blumberg JB, et al. Association Among Dietary Supplement Use, Nutrient Intake, and Mortality Among U.S. Adults: A Cohort Study. Ann Intern Med. 2019;170(9):604-613. doi:10.7326/M18-2478

134. Conlon MA, Bird AR. The impact of diet and lifestyle on gut microbiota and human health. Nutrients. 2014;7(1):17-44. Published 2014 Dec 24. doi:10.3390/nu7010017
135. Singh RK, Chang HW, Yan D, et al. Influence of diet on the gut microbiome and implications for human health. J Transl Med. 2017;15(1):73. Published 2017 Apr 8. doi:10.1186/s12967-017-1175-y
136. Thorburn AN, Macia L, Mackay CR. Diet, metabolites, and "western-lifestyle" inflammatory diseases. Immunity. 2014;40(6):833-842. doi:10.1016/j.immuni.2014.05.014
137. Sonnenburg JL, Bäckhed F. Diet-microbiota interactions as moderators of human metabolism. Nature. 2016;535(7610):56-64. doi:10.1038/nature18846
138. Baxter NT, Schmidt AW, Venkataraman A, Kim KS, Waldron C, Schmidt TM. Dynamics of Human Gut Microbiota and Short-Chain Fatty Acids in Response to Dietary Interventions with Three Fermentable Fibers. mBio. 2019;10(1):e02566-18. Published 2019 Jan 29. doi:10.1128/mBio.02566-18.
139. Johnson AJ, Vangay P, Al-Ghalith GA, et al. Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans. Cell Host Microbe. 2019;25(6):789-802.e5. doi:10.1016/j.chom.2019.05.005
140. Zhang P, Wang F, Hu J, Sorrentino R. Towards personalized medicine: leveraging patient similarity and drug similarity analytics. AMIA Jt Summits Transl Sci Proc. 2014;2014:132-136. Published 2014 Apr 7.
141. Zhang X, Chou J, Wang F. Integrative analysis of patient health records and neuroimages via memory-based graph convolutional network. In: Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM), Singapore. 2018;767-770. doi:10.48550/arXiv.1809.06018
142. Gottesman O, Kuivaniemi H, Tromp G, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med. 2013;15(10):761-771. doi:10.1038/gim.2013.72
143. Li L, Cheng WY, Glicksberg BS, et al. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. Sci Transl Med. 2015;7(311):311ra174. doi:10.1126/scitranslmed.aaa9364
144. Saxena R, Carnevale K, Sharma K. Digital Pathology and AI: A Paradigm Shift in Pathology Education. Journal of Population Therapeutics and Clinical Pharmacology, 2023;30(9):454-462. doi:10.53555/jptcp.v30i4.2672
145. Zhang N, Ju Z, Zuo T. Time for food: The impact of diet on gut microbiota and human health. Nutrition. 2018;51-52:80-85. doi:10.1016/j.nut.2017.12.005

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 8, Issue 09, ISSN No. 2455-2143, Pages 85-96
Published Online January 2024 in IJEAST (http://www.ijeast.com)

HOW INDIA ACHIEVES SPACE MISSION SUCCESS ON A SHOESTRING BUDGET: A LESSON IN ECONOMIC EFFICIENCY

Lubhani Kapoor
Department of Economics and Finance
Delhi Technological University

Dr. Sayak Bhattacharya
Department of Economics and Finance
Centre for National Policy Research

Abstract: **In the realm of space exploration, India has emerged as a true contender, defying conventional wisdom by achieving remarkable success on a shoestring budget. While global superpowers like the USA, Russia, and China pour massive resources into their space endeavors, India has showcased the art of achieving efficient and effective space missions without overspending. This unique approach to space economics not only highlights India's technical prowess but also underscores the importance of economic pragmatism in an arena historically marked by extravagant expenditures.**

I. INTRODUCTION

Space exploration plays a vital role in a country's development. It has spurred a growing commercial space sector, creating jobs and economic opportunities in fields related to space technologies like satellite communication, asteroid mining, and space tourism. Space assets are decisive for national security, encompassing communication, surveillance, and navigation. Furthermore, space exploration aids in tracking space debris, and potential threats and studying climate change, ensuring the safety of space assets. Additionally, it enables us to deflect objects that could impact Earth. It also enhances our knowledge of the cosmos, including the origins of stars, galaxies, and planets. Planetary science provides insights into geology, atmosphere, and the potential for harboring life or valuable resources. Moreover, it is imperative to develop advanced technology that is compact, energy-efficient, and inexpensive.

It is essential to consider the cost associated with space exploration. The space economy, which plays a significant role in the development of space, varies from country to country depending on various factors. The space economy represents a substantial and growing sector with the potential to contribute to economic growth and job creation. The commercial space industry is rapidly expanding with companies like SpaceX, Blue Origin, and numerous other startups entering the space market. The industry offers investment opportunities, job prospects, and potential returns for businesses and investors. Studying space economics helps researchers and the government to comprehend the scale of the overall growth of the country.

Studying space economics assists nations and organizations in navigating partnerships, agreements, and collaboration, in space endeavors, promoting peaceful cooperation in the cosmos. Understanding and assessing the economic incentives and mechanisms for space sustainability are vital to secure the long-term usability of Earth's orbital environment. Government policies and regulations shape the space industry. Understanding the evolving regulatory landscape, including funding and other policies, is important for comprehending the overall picture of the development taking place in this sector. It is indispensable to bring alterations in laws and policies to safeguard space assets and maintain global security along with keeping economic objectives.

Currently, there are 77 space agencies worldwide, with 16 of them possessing launch capabilities. Among these agencies, NASA(National Aeronautics and Space Administration) stands out as the most notable. Other important agencies are India- ISRO (Indian Space Research Organisation), Europe - ESA (European Space Agency), Canada -CSA (Canada Space Agency), Russia-Roscosmos (The Russian Federal Space Agency), China-CNSA (China National Space Administration), and Japan- JAXA (Japan Aerospace Exploration Agency). [1]

NASA was established on July 29, 1958, with a mission to pioneer the future in space exploration, scientific discovery, and aeronautic research. Its specific objectives include exploring the solar system and beyond, understanding and protecting our home planet, and developing innovative technologies. Notably, NASA is the largest player in the field of space technology boasting an annual budget exceeding $ 24.04 billion for 2030.

A few of the remarkable achievements of NASA include the successful landing of the Perseverance Mars rover on February 18, 2021. To continue its further research in the

same NASA will require at least two more projects currently being planned by NASA and the European Space Agency.
In addition to this, it launched its mega Moon rocket for the first time, sending uncrewed Orion spacecraft around the Moon. The Webb Space Telescope, another groundbreaking project, provided record-breaking new images from the cosmos. Additionally, NASA conducted humanity's first-ever planetary defense demonstration by successfully altering the course of an asteroid.
One of NASA's upcoming missions, Artemis 2, is scheduled for 2024 and will send the first astronauts around the Moon in nearly 50 years. This mission will involve launching four astronauts on a lunar fly-board and an Orion spacecraft using a space System rocket [2-4].
The Indian Space Research Organisation (ISRO) was established on August 15, 1969, through the merger of two government organizations. ISRO's primary objective is to develop indigenous space technology and applications for the benefit of society and the economy. Among ISRO's achievements are the successful launch of its PSLV and GSLV rockets, the Chandrayaan -3 mission, and the deployment of several satellites into orbit. ISRO launched Aditya L1, India's first space mission to study the Sun, marking the nation's foray into solar research. Moreover, India has plans to launch Shukruyaan to map the surface of Venus and Gaganyaan human space (manned) flight missions. Before this historic mission, ISRO had planned two unmanned missions as part of its preparations, underlining India's growing presence in space exploration and space economy.
Other important missions include XPoSat (X-ray Polarimeter Satellite), which represents India's first dedicated polarimetry mission, aiming to study various dynamics of bright astronomical X-ray sources in extreme conditions. ISRO is collaborating with NASA on the development of NISAR (NASA-ISRO SAR) a low Earth orbit (LEO) observatory. Its mission is to map the entire globe in just 12 days, discovering valuable insights.[5-8]
The European Space Agency (ESA) established in 1975 is another influential player in the field of space exploration. ESA is a multinational space organization currently composed of 23 member states, and it is open to any European Union country that meets specific criteria. Much like NASA, ESA has a strong track record of developing successful spacecraft and satellites.
One of the biggest collaborations between ESA and NASA was the launch of the James Webb Space Telescope. ESA provided the Arian 5 spacecraft for this groundbreaking mission, emphasizing international cooperation .[9]
Currently, ESA is actively involved in developing a Mars rover called ExoMars. This mission has ambitious objectives including the search for biosignatures of past Martian life and detailed study of the surface of Mars. Some of its upcoming missions include Plato, Earthcare, Flex, and many more.
The Canada Space Agency (CSA) was established in 1989 and has a rich history of collaboration with international partners in space exploration. CSA has been an active participant in numerous successful missions. One of its remarkable collaborations is with the International Space Station (ISS), where the renowned Canadarm, built by MDA, has played a crucial role in station operations. The industry is also employed in robotic exploration of Mars, demonstrating its capabilities in planetary exploration.
Several exciting mission of CSA is CASTOR, which has seen ongoing investments in critical technology development since early 2021, with an anticipated completion date in December 2023. This $2.25 million which is a huge investment encompasses work on various aspects, including telescope optical and structure design, focal plane and large are detectors, a fine steering misto, etc.[10]
ROSCOSMOS the Russian Federal Space Agency was established in 1992 following the dissolution of the Soviet Union. It has taken a leading role in developing Russia's satellite navigation system, known as GLONASS. One of the significant contributions to space exploration is its operation of the Baikonur Cosmodrome in Kazakhstan, from where the renowned Soyuz rocket is launched. For many years, NASA depended on the Soyuz to transport astronauts to the International Space Station (ISS), underlining the international collaboration in space endeavors.
The industry is also renowned for space tourism. The agency has enabled six space tourists to travel to space, with each trip costing an estimated fee of at least $20 million (USD). In August 2023, Roscoms initiated Luna 25 mission to the moon. The launch was conducted using Russia's own Soyuz -2.1 b rocket. However, the mission encountered a setback due to engine malfunction. The industry further aims for the Luna-Glob mission. It is also setting its sights on Venus. The industry is planning to launch an orbiter to explore and study Venus, marking its continued involvement in ambitious space missions .[11]
China National Space Administration (CNSA) founded in 1993 stands as one of the younger space agencies globally but has rapidly made significant strides in space exploration. China became the third country, following Russia and the USA to independently send a human into orbit with its spacecraft. China has seen great success in various areas including lunar exploration. The Chang'e lunar landers and rovers have significantly contributed to our understanding of the moon's surface.
China has ambitious plans concerning human spaceflight which include constructing three module space stations. The Tianhe core module, Launched in late April, is currently hosting its first crew, marking a substantial step forward in China's human space flight. CNSA has also planned to launch a near-Earth - asteroid sample return mission to small body 469219 Kamo'oalewa around 2025. It also aims to improve national civil space infrastructure, supporting ground facilities, and enhancing communication and broadcasting, navigation, and positioning facilities. The major aim behind all this is to

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 8, Issue 09, ISSN No. 2455-2143, Pages 85-96
Published Online January 2024 in IJEAST (http://www.ijeast.com)

boost economic development further solidifying the presence of China in space exploration.[12]
In 2003 the Japan Aerospace Exploration Agency (JAXA) was established through the merger of three Japanese organizations. JAXA's mission statement emphasizes its commitment ' to contribute to peaceful use and exploration of space, while also on Earth-related issues'. One of its recent missions Hayabusa-2 achieved significant success and is aimed to bring samples from the Ryugu asteroid. Hayabusa- 2 is now embarked on a journey to explore 2 more asteroids. It will provide live coverage of the launch of two important missions: the X-ray Imaging and Spectroscopy Mission (XRISM) and the smart Lander for Investigating Moon (SLIM) . [13]
The primary objective of the research paper is to analyze India's space exploration economy and budget profoundly and to draw insightful comparisons with major international space agencies. It is building on the preceding discussion about space agencies such as NASA, ESA, Roscosmos, JAXA, and CSA. This section will specifically focus on India's current space economy, anticipated developments, policy and regulatory shifts, overarching goals, and notable achievements within the sector. The paper will also provide a comparative examination of India's space economy and budgetary considerations in contrast to those of prominent space agencies worldwide. This analysis will shed light on the financial dummies of space exploration in India and its global standing.

II. ANTICIPATED DEVELOPMENTS POLICY AND REGULATORY SHIFTS OVERARCHING GOALS AND NOTABLE ACHIEVEMENTS

A. Role of the private sector

In The Indian government has taken major steps to include the private sector in the space industry. While the government has historically driven the space sector, the involvement of the private sector is inevitable for advancing satellite-based services and ground-based systems. Increased private sector involvement brings in new resources and talent, expanding the capabilities of the space industry. This will lead to an increase in planetary exploration. Moreover, it will all lead to better exploration of resources and space assets. Enhanced private sector participation can position the Indian industry as a key player in the global space economy, leading to substantial employment opportunities in the technology sector and making India a global technology powerhouse.
The steps taken by the government to include the private sector are: In the 2019 budget, New Space India Limited was established to mass-produce small satellite launch vehicles and polar satellite launch vehicles in collaboration with the private sector through technology transfer. The Draft National Policy 2020 aimed to foster public-private partnership in space research and exploration activities, granting private access to ISRO's infrastructure, technical resources, and data. The Union Cabinet in 2020 approved the creation of the Indian National Space Promotion and Authorisation Center (IN-Space) to serve as a single-window agency for space resources. Additionally, recently India's space agency unveiled a draft " Humans in Space Policy 2021" to enable the participation of non-traditional players in space activities.
But these steps also include various challenges: Regulatory complexity is one of them. Companies often need help with the multiplicity of approvals and the cumbersome process involving procedural ambiguity. Moreover, there is a lack of clear regulations, and access to "patient " capital, necessary for long-term space projects, is crucial. Also, if there is a conflict of interest, ISRO acting both as a regulator and operator may face conflicts of interest during dispute resolution with private players. Although IN-SPACE is intended to be independent, its connection with ISRO raises potential concerns regarding conflicts of interest.
It is important to note that a stable long-term policy framework is crucial for realizing the potential of India's space sector and achieving the goal of ' Aatmanirbhar Bharat'. The government is incessantly working on attracting investments from foreign resources and expanding collaboration with them indicating the sector's potential for growth and innovation [14].

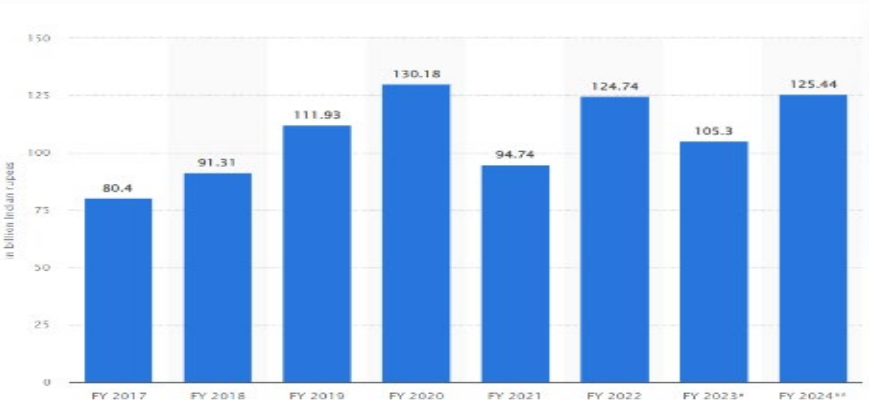


Fig. 1. Government expenditure on the space sector in India fro the financial year 2017- 2024 in billion Indian Rupees

B. Reforms

In June 2020, the Government of India (GOI) approved private enterprise participation in all phases of space activities and established the Indian National Space Promotion and Authorisation Centre
(IN-SPACE) within the India Space Research Organisation to support their independent space endeavors. Furthermore, India now boasts 368 private space firms, ranking fifth globally. The number of private company proposals in the space section increased by 30% in 2021.
Efforts to engage young entrepreneurs and students include the Atal Tinkering Lab (ATL) Space Challenge and the development of world-class learning facilities and space museums promoting STEM education. The government is also working to simplify the process of foreign investment to attract strong nations. The Indian Space Association (ISpA), launched in October 2021, aims to facilitate independent space activities, technology transfer from ISRO, and support for startups, MSMEs, and academia.
An MoU between the Department of Space (DoS) and NewSpace India Limited (NSIL) facilitates the transfer of ISRO technologies for commercialization with over 363 technologies transferred as of October 2021.
ISRO has adopted a demand-driven approach to space asset development and has opened its facilities for private sector use, allowing satellite testing at various centers and facilities, with IN-SPACe serving as the nodal agency for private engagement [15].
Facilities for testing, tracking, telemetry, launch pads, and laboratories, created by ISRO, would also enable the private space industry to climb the value chain. A mechanism has been created, where industries can approach IN-SPACe for utilizing the ISRO facilities [16].
A transition is taking place in the Indian space sector. Pawan Goenka the chairman of the Indian National Space Promotion Authorization Centre (IN-SPACe), said at a forum at the 74th International Astronautical Congress in Baku, on Oct 5 that ‘ we are moving from ISRO being the sole player in the space sector to the private sector taking on a more meaningful role’[17].

C. Laws and Regulations

India still does not have a space law to govern its activities in outer space. The government has its monopoly in India's space sector, which is led by the ISRO. The private sector has requested the government several times to open up the Indian space program. So businesses can invest to build India's space caliber.
The Indian Government introduced the Space Activities Bill 2017 to make a change in terms of space policies, but it has now been sent for further approval.
The Department of Space has entirely governed the space sector in India since its establishment in 1972 and before that, the Department of Atomic Energy took care. Outer space is no longer a domestic issue but has become an international issue. Hence the need for regulations is imperative. India now has big dreams to conquer with space technology. Henceforth, the laws and regulations should be accordingly established [18].
India is steadily emerging as a potential player in the global commercial space market, but several key legal and regulatory issues need attention. These include authorization, contacts, dispute resolution, licensing, data processing and distribution for earth observation services, insurance, certification of space technology, and legal matters related to launch services. Additionally, the revision of domestic laws is essential to address space law-related concerns.
There is a lack of legal protection for operators or the government in case of damages, intellectual property heights such as patents and data rights generated by ISRO, are not yet adequately addressed in space-related laws. Therefore, India’s domestic space legislation should encompass provisions for the peaceful use of outer space, ensuring holistic benefits for humanity.
India holds the 5th position in global space technology, signifying its valuable achievements in the field of space. Therefore space legalization is very crucial for its security. The current framework is shaped by the government of India’s policies and procedures, including the Remote sensing data policy of 2011, which removed restrictions on satellite data supply up to 1-meter resolution. Other policies include norms related to SATCOM and technology transfer by ISRO .[19]

D. Vision Mission Objectives

ISRO’s vision is to harness, sustain, and augment space technology for national development, while simultaneously advancing the frontiers of space, science research, and planetary exploration
ISRO’s mission encompasses various critical objectives for instance they are focused on designing and developing cutting-edge launch vehicles and related technologies to ensure reliable and cost-effective access to space. Additionally, they are computed to advance satellite technology, and associated systems for the earth are observation, communication, navigation, meteorology, and space science. Their communication program aims to establish robust telecommunication, television broadcasting, and developmental applications for effective national connectivity. Furthermore, their satellite-based remote sensing program plays a pivotal role in managing natural resources and monitoring the environment. Through space-based imagery promoting sustainable practices, they are dedicated to maintaining a space navigation system for precise and reliable navigation services across various applications. Leveraging space-based technologies. They aim to address societal challenges and enhance citizen’s quality of life. Their commitment to space, science, research, and planetary exploration drives ongoing research and development efforts expanding our understanding of the cosmos. Finally, they actively promote and authorize private firms to contribute to the global space market [20].

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 8, Issue 09, ISSN No. 2455-2143, Pages 85-96
Published Online January 2024 in IJEAST (http://www.ijeast.com)

E. Policy 2023

The government of India introduced the National Space Policy in 2023 aimed at harnessing India's distinctive expertise in the field of space exploration. This comprehensive policy serves as a dynamic framework to realize the government's vision for reforming the space sector. The key aspect of this policy is to grasp the attention of the private sector in space exploration. This will automatically boost the space economy and eventually the country's overall economy.

The guiding principles of the reforms include: firstly the policy seeks to establish a mechanism through which industries can engage with IN-SPACe for utilizing ISRO facilities. Additionally, it focuses on optimizing the utilization of space assets, including satellites and launch capacity, and defining accountability among various stakeholders. Importantly, the policy aims to create a favorable regulatory environment that provides equal opportunities for players within the Indian private sector. The policy aims to enhance India's space capabilities and promote a thriving commercial presence in space. The policy envisioned using space technology as a catalyst for technological advancements and driving benefits in related areas. It also emphasizes fostering international relations among all stakeholders.

Furthermore, under this policy, ISRO will concentrate on developing new technologies, systems, and research and development initiatives. NewSpace India Limited (NSIL) will carry out strategic activities in the space sector in a demand-driven mode moreover, IN-SPACe will interface with ISRO and non-government entities (NGEs), facilitating collaboration and engagement.

However, it is also important to understand the various challenges associated with this, including its capital-intensive nature, which may necessitate government investment in development. Furthermore, the limited allocation of funds for research and development poses obstacles. Additionally, the absence of space science programs in universities and colleges contributes to a shortage of skilled manpower in the sector [21].

F. National Security

In the 1960s when India initiated its space program, it was a developing nation grappling with limited resources and poverty. The essential focus was on leveraging space technology to advance social and economic development. However, over the past decade, there has been a change in perspective as development in other sectors proceeded: the pursuit of ambitious space exploration endeavors and an increased emphasis on space for national security purposes. This led to a completely new approach of India towards space and its economy. Space economy has now started accommodating these expenses as well. The transition was from morality and sovereignty-driven considerations to a more strategic and security-oriented perspective. Rather than being solely motivated by socio-economic development, India's space program now aligns itself with national security considerations.

Recognizing the evolving space security landscape, India is increasingly engaging with like-minded Partners. A notable avenue for cooperation is the Quadrilateral Security Dialogue. Or Quad. A crucial area where India can collaborate with Quad is enhancing Space Domain Awareness (SDA). SDA involves the capacity to monitor and track the space environment for potential threats, including both natural events and international attacks. [22-23]

India's space program has evolved from its humble beginnings driven by development goals to a multifaceted endeavor encompassing space exploration and national security imperatives. This transformation reflects the nation's pragmatism and changing priorities.

III. COMPARISON

A. INDIA VS CHINA

China's space capabilities have seen notable growth outpacing many nations in the global space race. In 2021, China made a significant mark by conducting an impressive 55 space missions, setting a record for the number of launches. This number far surpassed the total number of launches of the US and all other Asian countries combined, showcasing China's increasing prominence in space exploration.

In contrast, India's space activities in 2021 were marked by fewer launches, with just two missions. However, India's private space sector has seen remarkable success, standing in contrast to China's private space company, iSpace, which was incorporated in 2016 and has faced struggles. Despite a lower number of launches, India has been making strides in the space industry with a range of space capabilities.

India is positioning itself as a reliable alternative to companies like SpaceX, leveraging the geopolitical isolation of China and Russia with the hope of making lucrative business in the space sector leading to a rise in the overall economy. Concerns about Beijing's access to Western technology have led many satellite operators to avoid Chinese rockets boosting India's appeal. With ambitions to expand its role in the lucrative space industry, India is making steady progress.

When comparing space assets as of March 2020 China owned 13.6% of all earth-orbiting satellites, while India has a smaller share at 2.3%. China has been actively launching communication satellites with companies like GalaxySpace and Galactic Energy successfully placing satellites into low-earth orbit. On the other hand, India's launches have been primarily conducted by the commercial sector that is ISRO and NewSpace, with fewer launches. [24-25]

B. INDIA VS USA

The NASA budget represents less than half of the US space budget, with the rest being spent largely by the Department of Defense. Military interests have also played a significant role in the history of space activities. Between 1978 and 1986,

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 8, Issue 09, ISSN No. 2455-2143, Pages 85-96
Published Online January 2024 in IJEAST (http://www.ijeast.com)

about $ 10 billion of the $ 21 billion in sales and savings to the US industry, attributed to NASA's work, came from additional sales of commercial aircraft.

India's expenditure on space activities is remarkably smaller compared to NASA's budget. However, it's important to note that the Department of Space spent 15% of the central government's research budget in 2009-10, second only to the US which was 17 % in 2013, while European Union Countries allocated less than 5 % for space research. Given India's relatively low overall spending on research and development. This raises doubts about the inappropriate spending on space activities is the most acceptable and efficient allocation.

Narayan Prasad has highlighted critical and different concerns by stating that a dollar spent on NASA yielded $10 for the American economy however no such worthy trend could be noticed for ISRO. In the case of India understanding the relationship between expenditure and return is quite complex to determine. India's Mars mission in contrast cost only $ 74 million a fraction of the cost incurred by others, including NASA's $ 108 million budget for the Hollywood film " The Martian". Despite having a budget of 20 times smaller than NASA, ISRO has excelled in terms of efficiency, proficiency, resourcefulness, and cost-effectiveness.

While NASA and ISRO initially competed in the space race, they have now shifted towards collaboration, working together to advance space exploration for the benefit of humanity. This collaboration exemplifies the changing landscape of international space efforts, emphasizing the collective pursuit of knowledge and progress beyond national boundaries.[26-28]

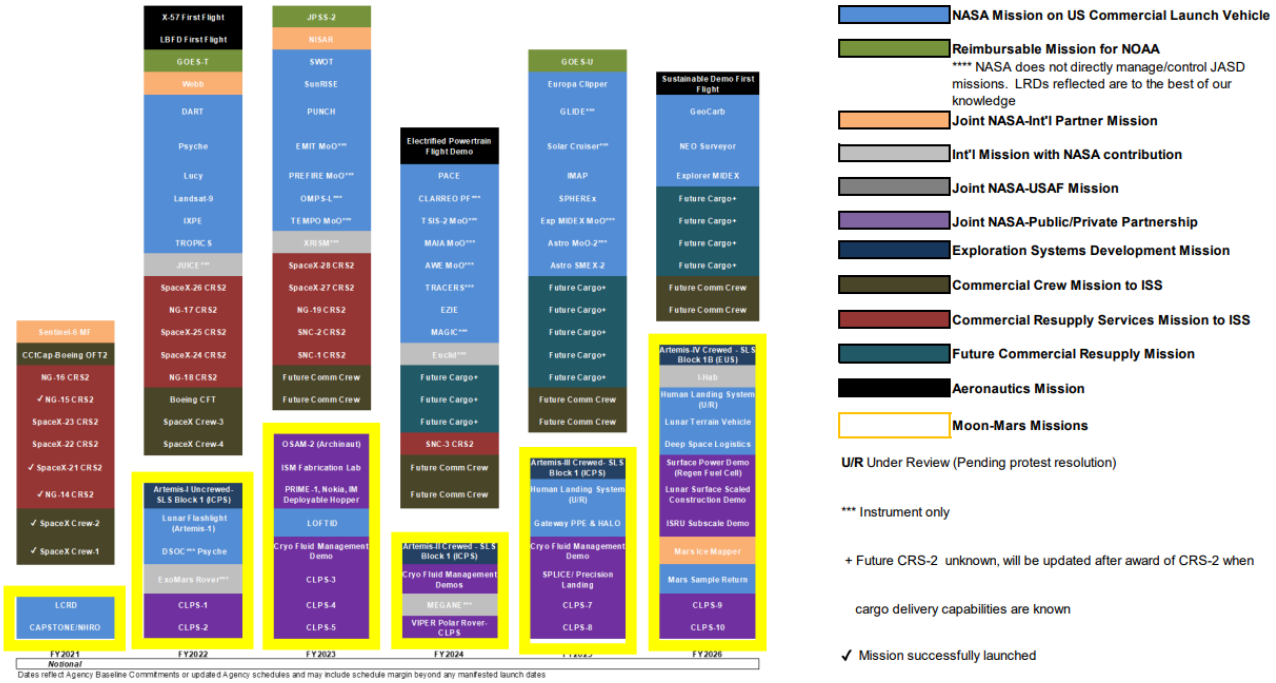


Fig. 2. NASA Mission planning manifest: FY2021 - 2026

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 8, Issue 09, ISSN No. 2455-2143, Pages 85-96
Published Online January 2024 in IJEAST (http://www.ijeast.com)

Budget Authority ($M)	FY 2020[1]	FY 2021[2]	FY 2022	FY 2023	FY 2024	FY 2025	FY 2026
Deep Space Exploration Systems	5,959.8	6,517.4	**6,880.4**	7,014.1	7,263.7	7,514.9	7,772.8
Exploration Systems Development	4,512.8	4,544.6	**4,483.7**	4,384.0	4,219.0	3,888.0	3,867.0
Exploration Research & Development	1,447.0	1,972.8	**2,396.7**	2,630.1	3,044.7	3,626.9	3,905.8
Space Technology	1,100.0	1,100.0	**1,425.0**	1,454.5	1,486.4	1,519.2	1,552.9
Space Operations	4,134.7	3,988.2	**4,017.4**	4,109.3	4,103.3	4,103.3	4,103.3
International Space Station	1,516.1	1,321.6	**1,327.6**	1,309.7	1,279.4	1,284.5	1,284.5
Space Transportation	1,746.2	1,872.9	**1,771.7**	1,827.1	1,849.0	1,843.7	1,843.7
Space and Flight Support	857.4	776.6	**817.0**	786.4	788.8	789.0	789.0
Commercial LEO Development	15.0	17.0	**101.1**	186.1	186.1	186.1	186.1
Science	7,143.1	7,300.8	**7,931.4**	8,095.6	8,272.9	8,455.7	8,643.4
Earth Science	1,971.8	2,000.0	**2,250.0**	2,343.5	2,398.3	2,573.0	2,702.3
Planetary Science	2,712.6	2,699.8	**3,200.0**	3,196.3	3,266.5	3,226.9	3,168.7
Astrophysics	1,306.2	1,356.2	**1,400.2**	1,461.8	1,491.5	1,512.3	1,594.1
Heliophysics	724.5	751.0	**796.7**	803.3	816.6	833.6	858.5
James Webb Space Telescope	423.0	414.7	**175.4**	172.5	172.0	172.0	172.0
Biological and Physical Sciences	5.0	79.1	**109.1**	118.1	128.0	137.9	147.8
Aeronautics	783.9	828.7	**914.8**	933.7	954.1	975.2	996.8
STEM Engagement	120.0	127.0	**147.0**	150.0	153.3	156.7	160.2
Safety, Security, and Mission Services	2,913.3	2,936.5	**3,049.2**	3,112.3	3,180.5	3,250.8	3,323.0
Mission Services & Capabilities	1,849.7	1,918.3	**2,028.8**	2,070.8	2,113.7	2,157.6	2,202.4
Engineering, Safety, & Operations	1,063.6	1,018.2	**1,020.4**	1,041.5	1,066.8	1,093.2	1,120.6
Construction & Envrmtl Compl Restoration	432.5	428.5	**390.3**	398.4	407.1	416.1	425.3
Construction of Facilities	357.8	370.4	**315.6**	322.2	329.3	336.7	344.2
Environmental Compliance and Restoration	74.7	58.1	**74.7**	76.2	77.8	79.4	81.1
Inspector General	41.7	44.2	**46.0**	47.0	48.0	49.1	50.2
NASA Total	22,629.0	23,271.3	24,801.5	25,314.9	25,869.3	26,441.0	27,027.9

1/ - FY 2020 reflects funding amounts specified in Public Law 116-93, Consolidated Appropriations Act, 2020, as adjusted in NASA's FY 2020 Initial Operating Plan, except Exploration Ground System Development and Exploration CoF. Table does not reflect emergency supplemental funding provided for NASA and included in the Safety, Security, and Mission Services account, as specified in Public Law 116-136, the Coronavirus Aid, Relief, and Economic Security Act, totaling $60.0 million.

2/ - FY 2021 reflects funding amounts specified in Public Law 116-260, Consolidated Appropriations Act, 2021, as adjusted by NASA's FY 2021 Initial Operating Plan.

Fig. 3. FY 2022 Budget request: Agency

C. INDIA VS RUSSIA

The race between Russia's and India's space economies encompasses elements of science, national pride, and the pursuit of financial gains. India is strategically positioning itself within the expanding space industry, taking advantage of the geopolitical isolation of China and Russia. Furthermore, India has cultivated closer ties with the US, Australia, Japan, and other regional powers, while offering more cost-effective launch services compared to others.

Russia boasts a rich history in space activities and its Soyuz rockets are renowned for their reliability, Years of under-investments have started revealing cracks in its capabilities. The war between Russia and Ukraine is also a failure regarding the space budget. However Russia's caliber in human spaceflight is unmatched, but this is a highly complex engineering challenge.

On the other hand, India's growing economy contrasts with Russia's ongoing economic contractions. Russia relies on technology from the first space race, while India is developing modern tools that reflect recent advancements. Moreover, India's receptiveness to international collaboration and commerce in space blisters its position. Due to war, Russia will require a lot of years to recover from the loss despite regaining access to global funding and high-tech components as the economic damage is huge.[29-33]

D. INDIA VS JAPAN

India and Japanese space agencies have jointly assessed collaboration in various space domains, such as earth observation, lunar missions, etc. This partnership comes from the strong economy of India which has grasped the interest of the Japanese corporate sector. The interest can also count on the expanding market of India, abundant Human Resources, and cost-effectiveness.[34]

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 8, Issue 09, ISSN No. 2455-2143, Pages 85-96
Published Online January 2024 in IJEAST (http://www.ijeast.com)

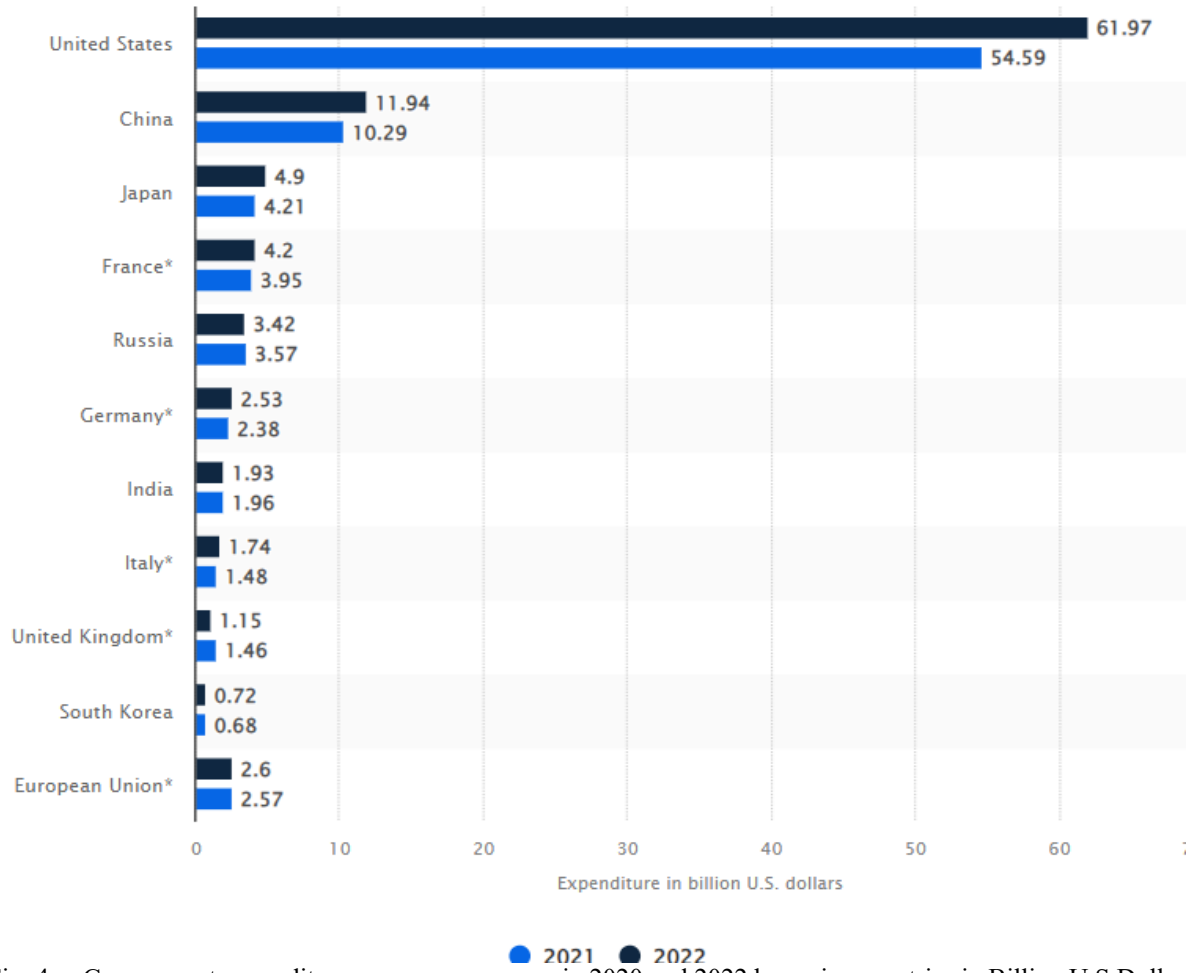


Fig. 4. Government expenditure on space programs in 2020 and 2022 by major countries in Billion U.S Dollars

India has traditionally not been able to spend a lot on its space endeavors. Today the government has aimed to spend a sufficient amount of money and make India a space power. India's space spending as a share of its GDP is relatively low. The US spends 0.28% of GDP, Russia is second at 0.15% while India is seventh in the world at 0.04% [35].

A recent EY report has put the global market size of the space sector at $ 447 billion in 2020 and is estimated to touch $ 600 billion by 2025.

According to the reports the Indian market size in 2020 amounted to $ 9.6 billion, which was 0.5 % of the GDP of the country.

The Indian government citing a study done by a global consulting firm told the parliament that the revenue generated by the the global space industry may increase to over $ 1 trillion by 2040.

Nekkanti, one of the research associates stated " By 2025, the satellite manufacturing segment will be the second fastest growing in the Indian space economy. Having proven our full stack technologies capabilities through our Thybolt Mission in November 2022, we are very excited to be a key player here,". [36-37]

IV. INDIA'S LOW-COST EDGE

India's success can be measured not only in its successful missions but also in its low-cost budget. India's budget-friendly space missions are the major highlight of ISRO's success. In 2024, Prime Minister Narendra Modi drew a dramatic comparison. He said India's Mars Orbiter Mission Mangalyaan at $ 74 million had cost less than the movie ' Gravity' which had cost nearly $ 100 million. NASA's similar Mars Mission, Maven, had cost nearly 10 times more.

India's leaders like Sarabhai had clear objectives for ISRO which is to work for the society, for the poor and not to compete with economically developed nations. From the very start, ISRO focused on the overall development of its countryman, and with this clear objective and working in a country which was badly ridden by poverty during those days their goal was to produce maximum benefits from minimum resources and efforts. Since 1947 India has struggled with

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 8, Issue 09, ISSN No. 2455-2143, Pages 85-96
Published Online January 2024 in IJEAST (http://www.ijeast.com)

paucity that has created a mindset against lavishness. The cost has to be kept in check always against lavish spending at NASA or the European Space Agency where costs can run away beyond initial estimates.

Then ISRO chairman Dr K Sivan had explained the economic nature of the space industry to TOI a few years ago." Simplifying the system, miniaturizing the complex big system, strict quality control, and maximizing the output from a product make our space missions frugal and cost-effective. We keep a strict vigil on every stage of development of a spacecraft or a rocket and therefore, we can avoid wastage of products, which helps us minimize the mission cost" he said.

In space engineering, where conditions are tough and costs of failure are high, it is not easy to adapt technology. It also involves more risk, but ISRO has been willing to take- and manage -that risk.

ISRO is well aware of optimizing the resources. Then ISRO chairman G . Madhavan Nair had told TOI " There are some tests that Americans would have done six times and we did only thrice. We scrutinized every parameter and optimized the tests. Yes, you may call it a calculated risk, but, touch wood, we have been successful with this approach so far. "

Moreover, ISRO has lighter payloads and heavier, slower, and less powerful rockets than the space agencies of the US and Russia. Russia's Luna, which failed to soft-land on the Moon a few days ago, took nearly one-fourth of the time to reach as compared to the time taken by Chandrayaan-3 to land.

India's low-cost manufacturing plays to the advantage of ISRO. Over time, ISRO has reduced imports and tried to domestically produce components as much as possible. This leads to major savings. This also leads to the development of the Indian economy by promoting the space market and creating employment. Participation of local industry in ISRO's projects for designing, manufacturing, and testing critical components and systems is well-known. L&T, Godrej Aerospace, Tata Consulting Engineers Limited, Mishra adha tu Nigam, and BHEL have supplied critical components for various missions including the recent ones at a fraction of the cost of imported components.

ISRO's surprisingly low costs are also attributed to the relatively low salaries of scientists. India's scientists have salaries of almost one-fifth of the scientists in the developed world.

The low cost is also associated with strategic partnership, ISRO actively engages in strategic collaboration with academic institutions, research organizations, and the private sector.[38-42]

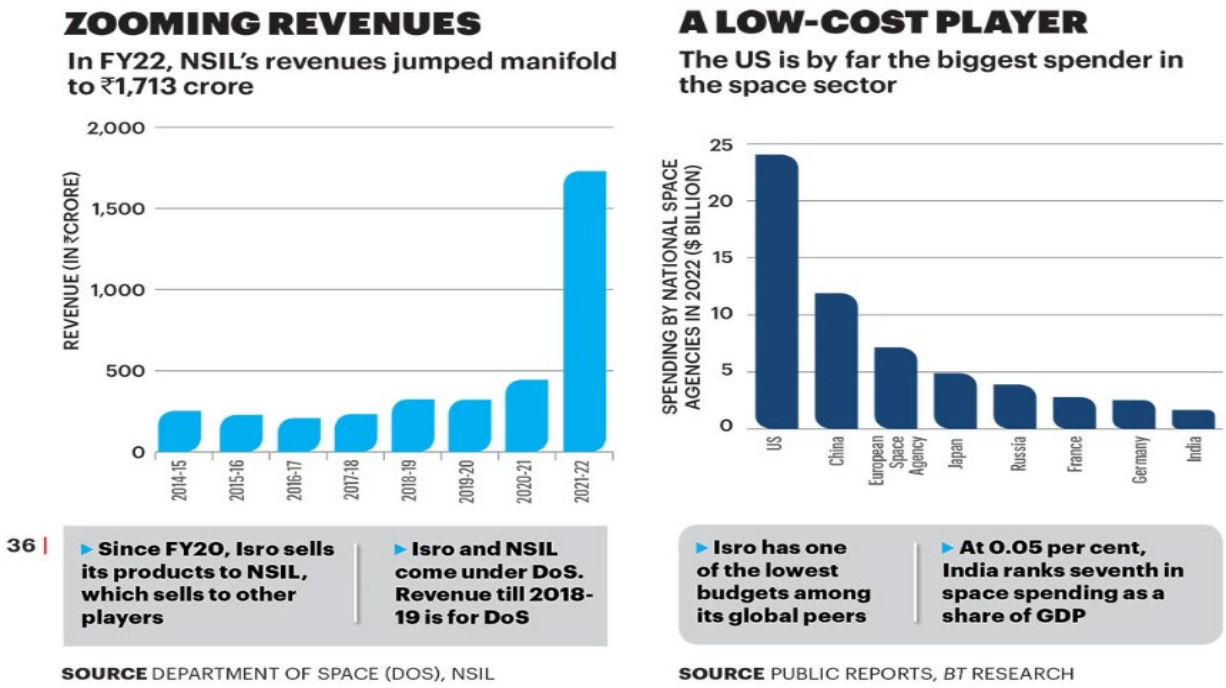


Fig. 5. Revenues vs Low-cost player

The satellite internet market is expected to touch a 27% share of the global space industry by 2040. A massive increase over its current share of less than 1 %, according to a Deloitte study. Moreover, satellite manufacturing, satellite launch, commercial human space flight, and commercial remote sensing are also likely to gain substantially. Crucially, the government spending component, which has been considerable all these years, is expected to fall from 27% in 2021 to less than 18% by 2040. ISRO will cede space to the private sector to drive future growth.

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 8, Issue 09, ISSN No. 2455-2143, Pages 85-96
Published Online January 2024 in IJEAST (http://www.ijeast.com)

ISRO has never been managed as a company as it always worked with a somewhat altruistic agenda. For example, the revenue earned from foreign and domestic satellite launches would earlier go into the Department of Space's (DOS) balance sheet. After NewSpace India Ltd (NSIL) was formed in 2019, it took care of all commercial activities after buying ready-made assets and services from ISRO. That money makes an entry in the balance sheet of the DOS, which in turn funds ISRO. NSIL is a profitable entity- in FY 22, it recorded a profit after tax of Rs 342.98 crore. Its revenue zoomed from Rs 321.77 crore in the financial year 2020 to Rs 1,731 crore in the financial year 2022.

ISRO has turned from a supply-based model to a demand-based model, after opening the Indian space ecosystem to private participation since 2000. It has already started helping with capacity building through the development of new technologies and capabilities, technology transfers to private players, and sharing its facilities with NSIL and Non-Government Private Entities.

Due to all these reasons the orbit economy booms, the cost of satellite launches continues to decline, and competition promises to get even more cut-throat, but the country remains in the reckoning.[43]

The low-cost budget has given India an edge in the global launch industry eventually attracting big economies. Under a pooch drive by our Prime minister, India is following NASA's lead in opening launch and other space businesses to private investment. India is aiming to increase its share of the global satellite launch market by fivefold within the next decade. While India can't compete with American private players like SpaceX, India's low cost will certainly help it compete with Russia and China who are ahead of India in launches.

But cutting down on cost leads to other consequences that cannot be ignored: It means India can collect less data and its missions are time-consuming. The entire process of space exploration is compromised, Chaitanya Girir, a fellow for space and ocean studies at Gateway House, stated.

Now that India has started opening up the space sector to private companies, ISRO may not remain solely dependent on government funding and thus would be able to spend more in the future. But more money would mean ISRO will chart out for itself and not it will lose its cost advantage .[44]

V. INTERNATIONAL COOPERATION

Over the years, ISRO's growing experience and technological advancements have expanded the opportunities for cooperation. International cooperation has been an integral part of India's space program from its inception. Initiatives such as the Thumba Equatorial Rocket Launching Station (TERLS), the Satellite Instructional Television Experiment (SITE), the Satellite Telecommunication Experiment Project (STEP), etc all included components of international cooperation.

ISRO actively fosters bilateral and multilateral relationships with space agencies and related organizations to enhance existing partnerships, tackle new scientific and technological challenges, refine space policies, and establish international frameworks for the peaceful use of outer space.

International cooperation is also vital to expand its budget and economy and achieve power. Developing countries try to take the help of ISRO to build their space technologies. ISRO's recent success has broadened the scope of international cooperation.

Few players in the space sector launch space vehicles at such a low cost as ISRO. Mc Dowell said, " And that is not Russia or China." ISRO has earned $ 94 million in foreign exchange from January 2018 to November 2022. From January 2018 to November 2022, ISRO has successfully launched 177 foreign satellites belonging to countries like Australia, Brazil, Canada, Colombia, Finland, France, Israel, Italy, Japan, Lithuania, Luxembourg, Malaysia, Netherlands, Republic of Korea, Singapore, Spain, Switzerland, UK, and the US. ISRO carried out two launches of its largest and heaviest rocket, in record time- on October 23, 2022, and on March 26, 2023. Both were commercial missions, where the LVM3 rocket ferried a total of 72 satellites of UK-based telecom firm One Web, thus earning ISRO a total revenue of more than Rs1100 cr or $ 137 million".[45]



Fig. 6. Value of exports in the space technology sector in India from FY 2015 - 2022. (In Billion Rupees).

VI. CONCLUSION

While India's model of cost-effective space exploration is commendable, it's essential to recognize that different nations have distinct goals, resources, and geopolitical considerations. Superpowers like the USA, Russia, and China often pursue broader strategic objectives, necessitating larger budgets for space programs. These countries also contribute to global technological advancements and scientific discoveries that benefit humanity as a whole.

India's success demonstrates that space exploration doesn't necessarily demand astronomical budgets. By embracing innovation, collaboration, and efficiency, countries can achieve significant milestones while containing costs. The economics of space travel are a delicate balance between ambition, necessity, and fiscal responsibility.

In a rapidly evolving landscape, as space becomes more accessible to a diverse range of nations and private entities, the lessons of India's economic approach could pave the way for a future where space exploration is focused on quantitative data mining and strategizing for future manned missions.[46]

VII. REFERENCE

[1]. https://rocketcrew.space/blog/space-agencies-around-the-world
[2]. https://www.planetary.org/spacemissions/perseverance#:~:text=Perseverance%20is%20a%20NASA%20Mars,18%2C%202021.
[3]. NASA. (2020). The Artemis Plan , (pg 9-11).
[4]. https://www.esa.int/Enabling_Support/Operations/Current_and_future_missions.
[5]. https://www.planetary.org/space-missions
[6]. https://www.isro.gov.in/XPoSat.html
[7]. Collins, P. (2000). Public Choice economics and space policy: Reliasing space tourism. Acta Astronautica, (Pg -12)
[8]. https://www.isro.gov.in/NISARSatellite.html
[9]. Robinson,P, and Westgaver, E.M. (1999). Eonomic Benefits from ESA Programmes . European Space Agency.
[10]. https://www.thehindu.com/sci-tech/science/what-next-for-isro-after-chandrayaan-3-mission/article67222472.ece
[11]. https://uk.news.yahoo.com/roscosmos-russias-space-agency-does-172928418.html#:~:text=What%20a
[12]. Jones Andrew. (2021). China outlines Space Plans to 2025. (pg1)
[13]. https://global.jaxa.jp/news/2023/#:~:text=September%202023&text=JAXA%20will%20provide%20the%20live,(H%2DIIA%20F47).&text=The%20broadcast%20date%20and%20time%20are%20subject%20to%20change
[14]. https://casca.ca/?p=19160
[15]. https://www.dhyeyaias.com/current-affairs/daily-current-affairs/private-player-involvement-in-space-sector
[16]. Jones Andrew. (2023). India hopes Reforms will make it a Global Space hub. (pg -1).
[17]. ISRO. (2022). Opening Up Space Sector.indd. (Pg 3-5)
https://spacenews.com/india-hopes-reforms-will-make-it-a-global-space-hub/
[18]. https://www.legalserviceindia.com/legal/article-9212-space-laws-in-india.html#:~:text=Indian%20Space%20Laws,Space%20Research%20Organisation%20(ISRO).
[19]. https://www.lexology.com/library/detail.aspx?g=4d90359c-b79e-40c3-8d2c-8c9fdf308484.
https://www.isro.gov.in/Vision-Mission-Objectives.html
[20]. https://byjus.com/free-ias-prep/indian-space-policy-2023-upsc-notes/
[21]. Pillai Rajagopalan . Rajeswari and Prasad Naryan. (2021). Space India 2.0 Commerce , Policy, Security and Governance Perspective
[22]. Rajagopalan Pillai Rajeswari . (2022). India's Space Priorities are Shifting towards National Security. (pg 1-3)
[23]. https://www.livemint.com/news/india/india-is-taking-on-china-in-447-billion-space-economy-11680666124594.html
[24]. https://www.deccanherald.com/science/space/analysis-isros-success-compared-to-pakistan-chinas-space-agencies-2657928
[25]. https://eightify.app/summary/technology-and-society/space-race-isro-vs-nasa-history-and-future-dhruv-rathee#:~:text=NASA%20has%20made%20major%20contributions,for%20India's%20socio%2Deconomic%20benefit.
[26]. https://seac-space.com/differences-between-nasa-and-isro/
[27]. Subramanian Kapil. (2016). The Problems with Measuring ISRO's Economic Impact. (pg 3)
[28]. Frenholz Tim. (2023). India is replacing Russia as a 21st-century space power, (pg 1)
[29]. https://www.reuters.com/science/indias-boom-russias-crunch-how-money-is-shaping-new-space-race-2023-08-22/
[30]. https://www.businessworld.in/article/India-Aims-At-447-Bn-Space-Economy-Amid-China-Russia-s-Isolation/05-04-2023-471724/
[31]. https://economictimes.indiatimes.com/news/science/indias-boom-russias-crunch-how-money-is-shaping-a-new-space-race/articleshow/102936728.cms?from=mdr

[32]. https://www.wionews.com/business-economy/how-the-money-is-influencing-a-new-space-race-with-indias-boom-and-russias-crunch-628286
[33]. https://www.isro.gov.in/InternationalCoOperation.html
[34]. B. R. Guruprasad. (2018). Understanding India's International Space Cooperation Endeavour. (pg 455-481)
[35]. https://economictimes.indiatimes.com/news/science/the-secret-behind-indias-budget-friendly-moonshots-how-isro-has-developed-its-low-cost-edge/articleshow/103018475.cms?from=mdr
[36]. https://telecom.economictimes.indiatimes.com/news/portal-in-portal/satcom/miles-to-go-indias-share-is-2-3-but-isro-is-now-an-end-to-end-player/102065518
[37]. https://economictimes.indiatimes.com/jobs/c-suite/indias-top-cxos-making-waves/articleshow/105012598.cms
[38]. https://medium.com/@chaiwithsai/isros-triumphs-business-lessons-to-learn-from-india-s-space-program-26e3a30e6f09
[39]. K.Kasturirangan and C. S. Shaijumon . (2021). Economics and Cost Benefit Analyisis of Indian Space Programme (DOI: 10.1007/978-981-33-6510-0 14)
[40]. Subrahmanian, K. (2016). The problems with measuring ISRO's economic impact. (pg 16-20).
[41]. Corrado Luisa, Cropper Maureen, Rao Akhil . (2023). Space Exploration and Economic growth: New Issues and Horizons. (Doi : 120 (43)e2221341120)
[42]. ISRO, Various years, Annual Reports of ISRO, Government of India
[43]. https://www.nextias.com/current-affairs/12-03-2021/india-japan-space-cooperation/https://jgu.s3.ap-south-1.amazonaws.com/jsia/Joseph+Punnen+-+Indo-Japan+Technology+Trade.pdf
[44]. Mani sunil, V. K. Dadhwal, C. S. Shaijumon. India's Space Economy 2011-12 to 2020-21:Its size and structure. (2022). (pg 6-8)

- [fig1.]https://www.statista.com/statistics/1220036/india-public-expenditure-on-space-sector/
- [fig2.]https://www.nasa.gov/wpcontent/uploads/2015/01/fy2022_budget_summary.pdf
- [fig.3.]https://www.nasa.gov/wp-content/uploads/2015/01/fy2022_budget_summary.pdf
- [fig4.]https://www.statista.com/statistics/745717/global-governmental-spending-on-space-programs-leading-countries/
- [fig. 5]https://www.businesstoday.in/magazine/deep-dive/story/the-isro-model-how-indias-space-agency-is-slowly-becoming-a-global-favourite-400231-2023-09-29.
- [fig.6]https://www.statista.com/statistics/1419641/india-space-technology-sector-export-value/

AIP Advances

RESEARCH ARTICLE | FEBRUARY 20 2024

Impact of external magnetic field on temperature anisotropy driven Weibel instability in a magnetized plasma

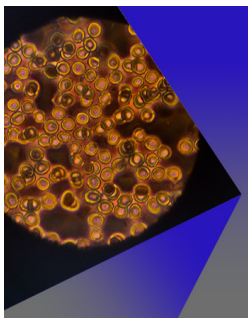
Twinkle Pahuja; Amit Kumar ; Himank Sagar; Ruby Gupta ; Jyotsna Sharma

Check for updates

AIP Advances 14, 025136 (2024)
https://doi.org/10.1063/5.0189656

View Online
Export Citation
CrossMark

26 February 2024 09:20:26



AIP Advances
Special Topic: Medical Applications of Nanoscience and Nanotechnology
Submit Today!
AIP Publishing

AIP Publishing

Impact of external magnetic field on temperature anisotropy driven Weibel instability in a magnetized plasma

Cite as: AIP Advances 14, 025136 (2024); doi: 10.1063/5.0189656
Submitted: 30 November 2023 • Accepted: 23 January 2024 •
Published Online: 20 February 2024

View Online Export Citation CrossMark

Twinkle Pahuja,[1] Amit Kumar,[2] Himank Sagar,[3] Ruby Gupta,[4] and Jyotsna Sharma[1,a)]

AFFILIATIONS

[1] Department of Physics, Amity School of Applied Sciences, Amity University Haryana, Manesar, Gurugram 122051, India
[2] Department of Physics, Modern Institute of Technology and Research Centre, Alwar, Rajasthan 301028, India
[3] Department of Applied Physics, Delhi Technological University, Shahbad Daulatpur, Bawana Road, Delhi 110042, India
[4] Department of Physics, Swami Shraddhanand College, University of Delhi, Alipur, Delhi 110036, India

[a)] **Author to whom correspondence should be addressed:** plasmajyoti@gmail.com

ABSTRACT

In this paper, we investigate the impact of an external magnetic field on the Weibel instability driven by temperature anisotropy in a magnetized plasma. The study provides the influence of electron temperature anisotropy, considering both cold as well as warm ions. We derive expressions for the growth rate Γ (s^{-1}) in each case to quantify the effect of some specific parameters on the instability. For example, the growth rate is plotted against the magnetic field for three distinct cases: considering only electrons, including cold ions, and incorporating warm ions. It is observed that the growth rate decreases as the applied magnetic field increases. Furthermore, the addition of both cold and warm ions to the plasma significantly influences the observed effects. It can be explained as follows: in the presence of cold ions, the growth rate is decreased with an increase in the wave number k_z (cm^{-1}). In addition, a decrease in the growth rate is observed with an increasing wave number in the case of warm ions. Therefore, it can be seen that the Weibel instability, propelled by thermal anisotropy, holds significant implications owing to its prevalence in both astrophysical and laboratory plasmas. The findings presented here are expected to contribute to the advancements in magnetic field generation and particle acceleration in various astrophysical phenomena, including gamma-ray bursts, relativistic collisionless shock fronts, radio supernovae, and pulsar winds.

© 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1063/5.0189656

26 February 2024 09:20:26

I. INTRODUCTION

In plasma physics, the mystery surrounding the mechanism of magnetic field creation in various plasmas, ranging from terrestrial to planetary and cosmological, has been a significant puzzle in the field. Weibel[1] proposed the Weibel instability in 1959 as a key process in amplifying magnetic fields and initiating self-seeding in plasmas. It can be observed from the literature that the instability[1–25] arises from the self-organization of minute currents within the static anisotropic plasmas. Here, anisotropy pertains to the variations in electron temperatures (or EVDs, or electron velocity distributions) across different spatial directions within the plasma. The Weibel instability, induced by temperature anisotropy, plays a crucial role in various astrophysical phenomena, such as laser fusion experiments and dense quantum plasmas. It serves as a dependable mechanism for generating magnetic fields in these diverse scenarios. The WI creates a collision-less shock by converting the kinetic energy (K.E.) of streaming electrons into the magnetic energy. Within astrophysical systems, the magnetic fields essential for the formation of collision-less shocks may either be pre-existing, as observed in phenomena such as supernova remnants or early galaxies,[2] or they can be self-generated, as exemplified in the context of gamma-ray bursts (GRBs).[3] The magnetic fields in GRB outflows possess greater strength compared to those achievable through the GRB progenitor or by having mismatched temperature and density gradients.[4,5] It has been understood for long that instabilities,[26–52] even in the lack of seed fields, can produce powerful magnetic fields. Huntington *et al.*[6] performed an experimental observation of the formation of a

© Author(s) 2024

magnetic field in interpenetrating plasma flows using Weibel instability. Here, they discovered that one contender for producing strong enough fields to produce a collision-less shock is the Weibel instability. In laser-driven laboratory experiments, they employed the proton probe to directly photograph the electromagnetic (EM) fields. For two models, (i) relativistic electron beam (REB) and (ii) relativistic monoenergetic beam (RMB), Zaki[7] investigated the impact of collisions inside the plasma on the instability growth rate of WI and discovered that the growth rate of WI is higher for REB as compared to RMB. Achterberg and Wiersma[8] addressed the formation of magnetic fields in ultra-relativistic shocks linked with (GRBs) using the linear theory of WI inside a plasma system driven via ultra-relativistic beams (URBs). A counter-streaming plasma system with temperature anisotropies was studied by Lazar *et al.*[9] in terms of its dispersion relations, characteristics, and stability. The growth rates were discovered to be less than those attained for Maxwellian plasmas. Borodachev and Koloniets[10] adopted particle-in-cell (PIC) numerical simulation to explain the usual Weibel instability picture. Here, the authors were able to track the dynamic remodeling of current filaments by obtaining the dependence of the fundamental properties of the temperature anisotropy parameters. Ryutov *et al.*[11] presented an explanation of the collisional effects on the ion (WI) for two counter-propagating plasma streams. They have found that the growth rate becomes much lower than that resulting from a collision-less model as dispersion relations for the situation of strong intra stream collisions are calculated. After researching the effects of thermal circumstances on Weibel instability, Mahdavi and Khanzadeh[12] discovered that as the collision rate rises, Weibel instability's growth rate decreases. Aman-ur-Rehman *et al.*[13] investigated the impact of collisions on WI with anisotropic electron distribution using analytical analysis, and they discovered that the growth rate γ (s^{-1}) of collisional plasmas turns negative if the frequency of electron-ion collisions exceeds the growth rate of the WI for a collision-less plasma. Grassi *et al.*[14] investigated WI induced via two relativistic symmetric counter-streaming electron plasmas under a uniform and constant external magnetic field, which is aligned along the plasma flows. Here, non-linear stages of the instability are examined using Particle-In-Cell (PIC) simulations. Silva *et al.*[15] provided an explanation of anisotropy using bi-Maxwellian anisotropy. Here, they observed that the growth rate wave vector changes their shape for non-Maxwellian distribution functions. The experimental observations on the generation of magnetic fields inside non-relativistic plasmas by using Ti :sapphire laser pulses (0.8 μm) were measured by Zhang *et al.*[16] Additionally, Zhang *et al.*[17] identified the phenomenon of own magnetic field mapping caused by heat. Here, they got the image the deflections of a REB with a (10^{-15} s) picosecond temporal period to capture the 2D development of the magnetic field connected to the WI and demonstrated the recorded k-resolved growth rates of the instability. Recently, Kumar *et al.*[18] have studied the Electromagnetic (EM) Weibel instability in spatial anisotropic electron-ion plasmas using fluid theory, and the expressions of growth rates are derived. Additionally, Kumar *et al.*[19] investigated the temperature anisotropy-driven WI in carbon plasmas, where the expression of the growth rate is derived in the manifestation of carbon and hydrogen ions. This temperature anisotropy-driven EM instability has been detected or identified in laser-generated plasmas,[20] interplanetary plasmas,[21] experimental plasmas,[22,23] the microwave disintegration of dilute gases,[24] as well as laboratory plasmas. In the fast ignition (FI) scenario, this instability plays a unique role and is crucial for the transfer of energy.[25,26] One of the most major causes for the generation of magnetic fields in space,[27,28] supernovas,[29] and galactic cosmic rays is responsible for this instability. This paper investigates the influence of an external magnetic field on temperature anisotropy-driven Weibel instability in a magnetized plasma by solving a unified set of Maxwell's equations for three distinct cases when electrons, cold ions, and warm ions are present.

Finally, this article is structured as follows: Sec. II A provides the details of the numerical expression of the growth rate in the existence of electron temperature anisotropy. In Sec. II B, the expressions of the growth rate in the existence of cold ions and electron temperature anisotropy are derived. However, the expression of the growth rate in the existence of warm ions is derived in Sec. II C. Result and discussion parts are provided in Sec. III, whereas Sec. IV presents the conclusion part.

II. INSTABILITY ANALYSIS FOR WEIBEL INSTABILITY

Let us assume a plasma system comprising of plasma electrons having density n^{0e} and ions having density n^{0i}. The mass, charge, and temperature of ions and electrons are given as $(m^i, e, \vec{T^i})$ and $(m^e, -e, \vec{T^e})$, respectively. For ions and electrons, the temperature (having various x and z directions) is expressed as $\vec{T^i} = T^{xi}\,\hat{x} + T^{zi}\,\hat{z}$ and $\vec{T^e} = T^{ex}\,\hat{x} + T^{ez}\,\hat{z}$. An exterior magnetic field ($B_s\hat{z}$) is applied in the direction of the z axis and plasma electron stream at a velocity $\vec{v_e} = v_{tx}\hat{x}$ inside the plasma. Consider an EM wave transmitting along the z direction through this complex plasma system having an electric field,

$$\vec{E}_1 = A\, e^{-i(\omega t - k_z z)}\,\hat{x} \tag{1}$$

and

$$\vec{B}_1 = \frac{c\,k_z\,E_1}{\omega}\,\hat{y}. \tag{2}$$

For electrons and ions, the equation of motion and continuity is expressed as

$$m^e\left[\frac{\partial \vec{v}_e}{\partial t} + (\vec{v}_e\cdot\nabla)\vec{v}_e\right] = -e\vec{E}_1 - \frac{e}{c}[\vec{v}_e\times(\vec{B}_s+\vec{B}_1)] - \vec{T^e}\frac{\nabla n^{1e}}{n^{0e}}, \tag{3}$$

$$m^i\left[\frac{\partial \vec{v}_i}{\partial t} + (\vec{v}_i\cdot\nabla)\vec{v}_i\right] = e\vec{E}_1 + \frac{e}{c}[\vec{v}_i\times(\vec{B}_s+\vec{B}_1)] - \vec{T^i}\frac{\nabla n^{1i}}{n^{0i}}, \tag{4}$$

$$\left[\frac{\partial n^e}{\partial t} + \nabla\cdot(n^e\vec{v}_e)\right] = 0, \tag{5}$$

and

$$\left[\frac{\partial n^i}{\partial t} + \nabla\cdot(n^i\vec{v}_i)\right] = 0. \tag{6}$$

26 February 2024 09:20:26

© Author(s) 2024

The linearization of Eq. (3) gives the x, y, and z component of perturbed electron velocities as

$$v^{e\,1^x} = -\frac{e\,E_1\,i\,\omega}{m^e\,(\omega^2-\omega_{ce}^2)}, \tag{7}$$

$$v^{e\,1^y} = -\frac{v^{e\,1^x}\,\omega_{ce}}{i\,\omega}, \tag{8}$$

and

$$v^{e\,1^z} = \frac{e\,v_{tx}\,k_z\,E_1}{m^e\,i\,\omega^2} + \frac{T^{ez}\,k_z\,n^{1e}}{m^e\,n^{0e}\,i\,\omega}. \tag{9}$$

The axial and perpendicular components of the perturbed electron velocities are v^{e1z} and v^{e1x}, respectively.

Using the linearization procedure in Eq. (5), we have the value of electron perturbed density as

$$n^{1e} = \frac{n^{0e}e\,k_z^2 v_{tx}\,E_1}{m^e\,i\,\omega\,(\omega^2-k_z^2\,v_{tz}^2)}. \tag{10}$$

We determine the expression of perturbed current density for plasma electrons that are streaming along the x direction using Eqs. (7) and (10) given as

$$J_{e1x} = -\frac{n^{0e}\,e^2\,E_1}{m^e\,i\,\omega}\left[\frac{\omega^2}{\omega^2-\omega_{ce}^2} + \frac{k_z^2\,v_{tx}^2}{(\omega^2-k_z^2\,v_{t\,z}^2)}\right], \tag{11}$$

where $v_{tx} = \left(\frac{T^{ex}}{m^e}\right)^{1/2}$ and $v_{tz} = \left(\frac{T^{ez}}{m^e}\right)^{1/2}$ represent the electron thermal velocity along x and z directions.

Similar to this, if plasma ions are streaming along the x direction, Eq. (11) can likewise be used to compute the current density of ions by exchanging $n^{0e}, m^e, v_{tx}^2, v_{tz}^2$ to $n^{0i}, m^i, v_{txi}^2, v_{tz\,i}^2$, respectively, as

$$J_{i1x} = -\frac{n^{0i}\,e^2\,E_1}{m^i\,i\,\omega}\left[\frac{\omega^2}{\omega^2-\omega_{ci}^2} + \frac{k_z^2\,v_{txi}^2}{(\omega^2-k_z^2\,v_{tzi}^2)}\right]. \tag{12}$$

A. The impact of the external magnetic field on Weibel instability in the existence of electron temperature anisotropy

Using Maxwell's equations, the wave equation may be expressed as follows (when there are no ions present and electrons are streaming in the direction of the x axis):

$$\nabla^2\vec{E}_1 - \nabla(\nabla\cdot\vec{E}_1) + \frac{\omega^2}{c^2}\vec{E}_1 = -\frac{4\pi i\omega}{c^2}J_{e1x}. \tag{13}$$

Using the value of J_{e1x} from Eq. (11) into Eq. (13), we get $-k_z^2E_1 + \frac{\omega^2E_1}{c^2} = \frac{4\pi\,i\,\omega\,n^{0e}\,e^2\,E_1}{c^2\,m^e\,i\,\omega}\left[\frac{\omega^2}{(\omega^2-\omega_{ce}^2)} + \frac{k_z^2\,v_{tx}^2}{(\omega^2-k_z^2\,v_{tz}^2)}\right]$.

On simplifying the above equation, we get

$$\omega^2 - k_z^2c^2 + \frac{\omega_{pe}^2\omega^2}{\omega_{ce}^2} = \frac{\omega_{pe}^2k_z^2v_{tx}^2}{(\omega^2-k_z^2v_{tz}^2)} \tag{14}$$

or

$$\omega^4\left(1+\frac{\omega_{pe}^2}{\omega_{ce}^2}\right) - \omega^2\left(k_z^2v_{tz}^2 + k_z^2c^2 + \frac{\omega_{pe}^2k_z^2v_{tz}^2}{\omega_{ce}^2}\right) + k_z^4v_{tz}^2c^2 - \omega_{pe}^2k_z^2v_{tx}^2 = 0. \tag{15}$$

The root of Eq. (15) providing the growth rate is given as

$$\omega^2 = -\left[\frac{\frac{\omega_{pe}^2v_{tx}^2}{c^2} - k_z^2v_{tz}^2}{\omega_{PR}^2k_z^4}\right]. \tag{16}$$

The growth rate can be expressed as $\Gamma = \mathrm{Im}(\omega)$, which results in

$$\Gamma = \left[\frac{\frac{\omega_{pe}^2v_{tx}^2}{c^2} - k_z^2v_{tz}^2}{\omega_{PR}^2k_z^4}\right]^{1/2}, \tag{17}$$

where $\omega_{PR}^2 = \frac{k_z^2c^2}{\left(1+\frac{\omega_{pe}^2}{\omega_{ce}^2}\right)}$ and $\omega_{pe} = \left(\frac{4\pi n^{0e}e^2}{m^e}\right)^{1/2}$ signify the electron plasma frequency.

B. The impact of the external magnetic field on Weibel instability in the existence of cold ions

In this instance, the overall perturbed current density is equal to the sum of the current densities of the plasma electrons and ions, i.e.,

$$J_{1x} = J_{e1x} + J_{i1x},$$

$$J_{1x} = -\frac{n^{0e}\,e^2\,E_1}{m^e\,i\,\omega}\left[\frac{\omega^2}{\omega^2-\omega_{ce}^2} + \frac{k_z^2\,v_{tx}^2}{(\omega^2-k_z^2\,v_{tz}^2)}\right] - \frac{n^{0i}\,e^2\,E_1}{m^i\,i\,\omega}\left[\frac{\omega^2}{\omega^2-\omega_{ci}^2} + \frac{k_z^2\,v_{txi}^2}{(\omega^2-k_z^2\,v_{tzi}^2)}\right]. \tag{18}$$

Consequently, using Maxwell's equations, the wave equation in the existence of cold ions may be expressed as

$$\nabla^2\vec{E}_1 - \nabla(\nabla\cdot\vec{E}_1) + \frac{\omega^2}{c^2}\vec{E}_1 = -\frac{4\pi i\omega}{c^2}J_{1x}, \tag{19}$$

$$-k_z^2E_1 + \frac{\omega^2E_1}{c^2} = \frac{4\,\pi\,i\,\omega\,n^{0e}\,e^2\,E_1}{c^2\,m^e\,i\,\omega}\left[\frac{\omega^2}{\omega^2-\omega_{ce}^2} + \frac{k_z^2\,v_{tx}^2}{(\omega^2-k_z^2\,v_{tz}^2)}\right] + \frac{4\pi\,i\,\omega\,n^{0i}\,e^2\,E_1}{c^2\,m^i\,i\,\omega}\left[\frac{\omega^2}{\omega^2-\omega_{ci}^2} + \frac{k_z^2\,v_{txi}^2}{(\omega^2-k_z^2\,v_{tzi}^2)}\right].$$

When we simplify the wave equation given above, we get

$$-k_z^2c^2 + \omega^2 = \frac{\omega_{pi}^2\omega^2}{\omega^2-\omega_{ci}^2} + \frac{\omega_{pi}^2k_z^2v_{txi}^2}{(\omega^2-k_z^2v_{tzi}^2)} + \frac{\omega_{pe}^2\omega^2}{\omega^2-\omega_{ce}^2} + \frac{\omega_{pe}^2k_z^2v_{tx}^2}{(\omega^2-k_z^2v_{tz}^2)}. \tag{20}$$

We consider a scenario in which the ions are cold along the x axis, causing the streaming velocity to be extremely low and almost nil, i.e., $v_{txi} \approx 0$.

26 February 2024 09:20:26

© Author(s) 2024

Then, Eq. (20) can be revised as

$$\omega^4\left(1+\frac{\omega_{pe}^2}{\omega_{ce}^2}+\frac{\omega_{pi}^2}{\omega_{ci}^2}\right)-\omega^2k_z^2c^2=\omega_{pe}^2k_z^2v_{tx}^2. \tag{21}$$

The result of simplifying the aforementioned equation is given by

$$\omega^2=-\left[\frac{\omega_{pe}^2\,v_{tx}^2}{\omega_{PQ}^2}\right], \tag{22}$$

which results in the growth rate as

$$\Gamma=\left[\frac{\omega_{pe}^2\,v_{tx}^2}{\omega_{PQ}^2}\right]^{1/2}, \tag{23}$$

where $\omega_{PQ}^2=\frac{k_z^2c^2}{1+\frac{\omega_{pe}^2}{\omega_{ce}^2}+\frac{\omega_{pi}^2}{\omega_{ci}^2}}$.

C. The impact of the external magnetic field on Weibel instability in the existence of warm ions

In the presence of warm ions, Eq. (20) can be rewritten as

$$\omega^2-k_z^2c^2=-\frac{\omega_{pi}^2\omega^2}{\omega_{ci}^2}-\frac{\omega_{pe}^2\omega^2}{\omega_{ce}^2}+\frac{\omega_{pi}^2k_z^2v_{txi}^2}{\omega^2-k_z^2v_{tzi}^2}+\frac{\omega_{pe}^2k_z^2v_{tx}^2}{\omega^2-k_z^2v_{tz}^2}. \tag{24}$$

As we know, $v_{tzi}^2\ll v_{tz}^2$, and for $\frac{\omega}{k_z}\gg v_{tz}>v_{tzi}$, Eq. (24) can be simplified as

$$\omega^4-\omega^2\omega_{PQ}^2-\frac{\omega_{PQ}^2\left(\omega_{pe}^2\,v_{tx}^2+\omega_{pi}^2\,v_{txi}^2\right)}{c^2} \tag{25}$$

or

$$\omega^2=-\left[\frac{\omega_{pe}^2\,v_{tx}^2+\omega_{pi}^2\,v_{txi}^2}{\omega_{PQ}^2\,c^2}\right].$$

Therefore, the growth rate can be written as

$$\Gamma=\left[\frac{\omega_{pe}^2\,v_{tx}^2+\omega_{pi}^2\,v_{txi}^2}{\omega_{PQ}^2\,c^2}\right]^{1/2}. \tag{26}$$

III. RESULTS AND DISCUSSION

We have theoretically examined the impact of an exterior magnetic field on the Weibel instability using fluid theory. The numerical parameters used in the present study are plasma ion density $n^{0i}=10^9$ cm^{-3}, electron plasma density $n^{0e}=10^9$ cm^{-3}, and magnetic field $B_s=3000$ G. Utilizing Eq. (17), we have plotted the growth rate against the applied magnetic field in Fig. 1. It is evident from the observation that as the value of the applied magnetic field increases, the growth rate decreases. This behavior is attributed to the stabilizing effect of an increased external magnetic field on the plasma system. Furthermore, employing Eq. (17), Fig. 2 is constructed, illustrating the variation of the growth rate Γ (s^{-1}) with the parallel wave number k_z (cm^{-1}) for different values of the applied external magnetic field $B_s=300$ G and $B_s=3000$ G. In Fig. 2, it is evident that as the wave number increases, the growth rate of the destabilizing mode decreases. This observation indicates that electrons and ions in the plasma become excited in the direction of the magnetic field, thereby lowering the energy of the electromagnetic plasma along the path of the self-generated magnetic field. Consequently, the growth rate of

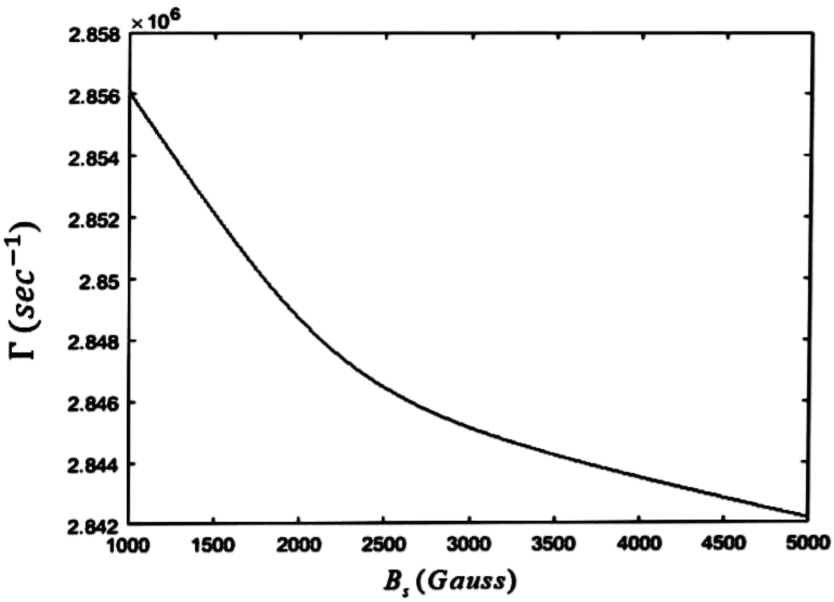


FIG. 1. The growth rate $\Gamma(s^{-1})$ vs applied magnetic field $B_s(G)$.

26 February 2024 09:20:26

© Author(s) 2024

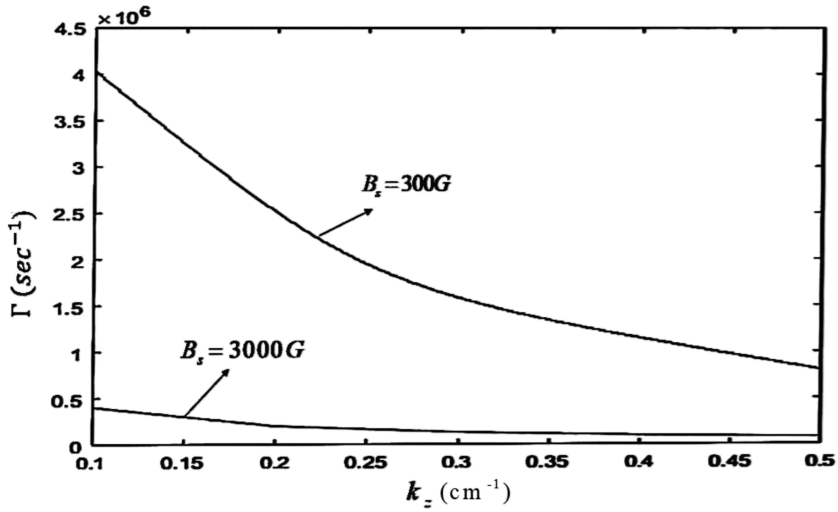


FIG. 2. The growth rate $\Gamma(s^{-1})$ vs wave number $k_z(cm^{-1})$ for different values of the applied magnetic field B_s = 300 (G) and B_s = 3000 (G), respectively, in the existence of electron temperature anisotropy.

the instability diminishes in the direction of the self-generated magnetic field. This reduction in the growth rate is also reported with the parallel wave number $k_z(cm^{-1})$.

In Fig. 3, the growth rate vs wave number $k_z(cm^{-1})$ is plotted for different values of temperature of electrons along the x direction with T^{ex} = 0.4 eV, T^{ex} = 0.6 eV, and T^{ex} = 0.8 eV in the existence of electron temperature anisotropy. Here, we observed that the instability increases by increasing the value of temperature anisotropy. This is due to the fact that as the temperature of electrons increases, the electron thermal velocity along the x direction also increases. Hence, the instability is increased. We have displayed Fig. 4 using Eq. (23) to show the growth rate vs parallel wave number k_z (cm^{-1}), when only cold ions are present inside the plasma, i.e., v_{txi} = 0, so the growth rate decreases with the rise in the value of the magnetic field. Figure 5 shows the growth rate $\Gamma(s^{-1})$ vs parallel wave number $k_z(cm^{-1})$ for different values of the

26 February 2024 09:20:26

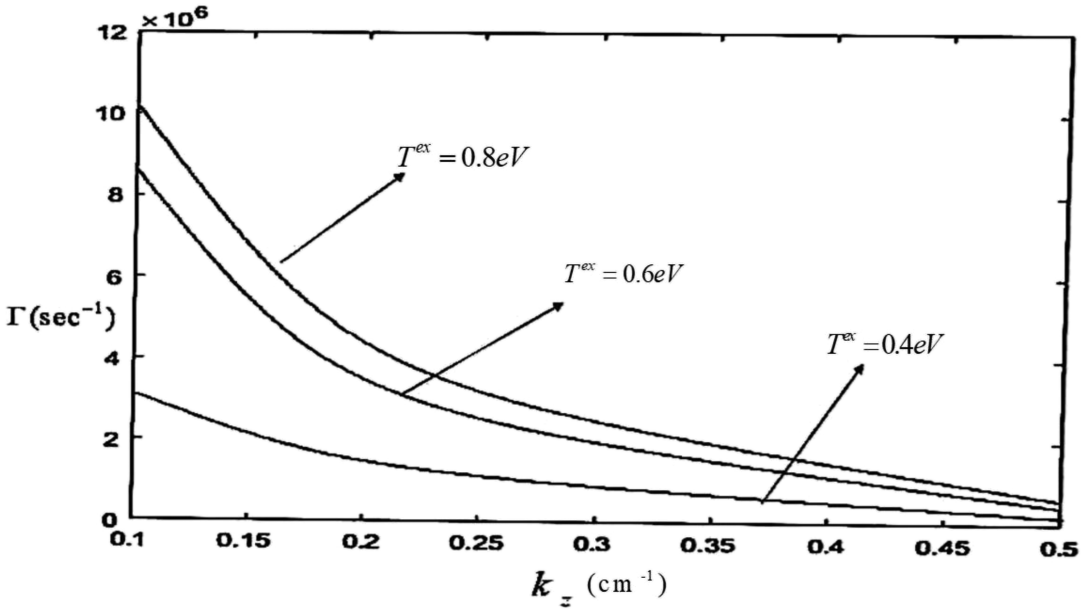


FIG. 3. The growth rate $\Gamma(s^{-1})$ vs wave number $k_z(cm^{-1})$ for different values of temperature of electrons along the x direction in the existence of electrons.

© Author(s) 2024

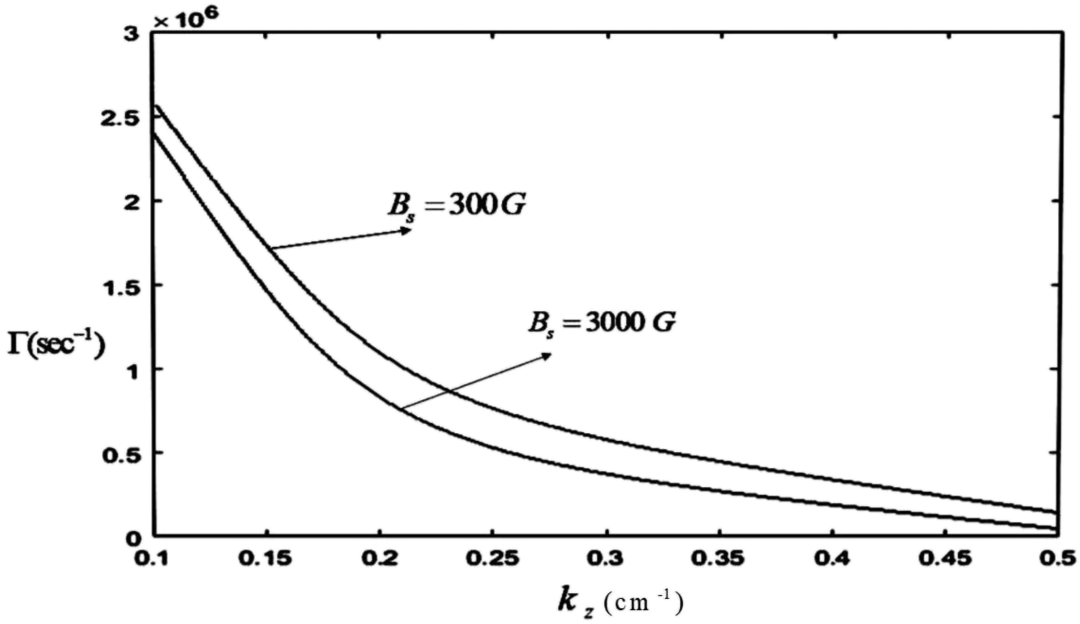


FIG. 4. The growth rate $\Gamma(s^{-1})$ vs wave number $k_z(cm^{-1})$ for different values of the applied magnetic field B_s = 300 (G) and B_s = 3000 (G), respectively, in the existence of cold ions.

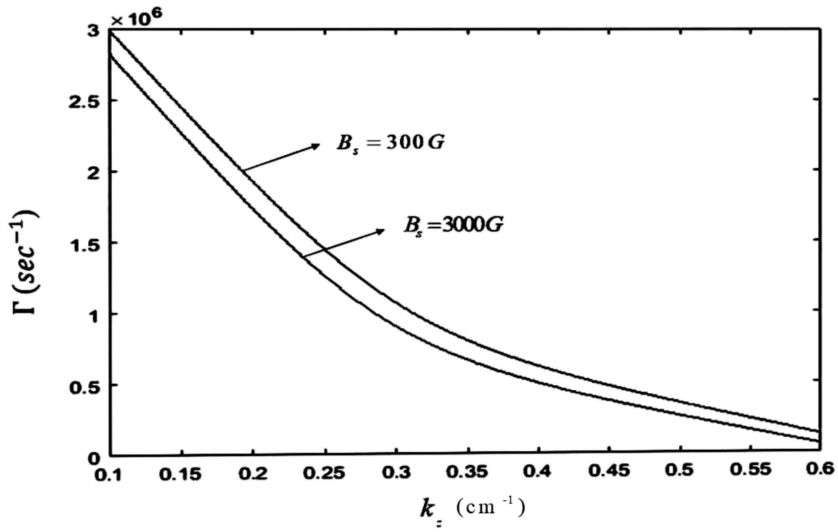


FIG. 5. The growth rate $\Gamma(s^{-1})$ vs wave number $k_z(cm^{-1})$ for different values of the applied magnetic field B_s = 300 (G) and B_s = 3000 (G), respectively, in the existence of warm ions.

applied magnetic field B_s = 300 G and B_s = 3000 G in the manifestation of warm ions. It is observed from Fig. 5 that the growth rate value is larger in the case of warms ions as compared to the cold ion case. Therefore, the growth rate of the Weibel instability decreases with the increase in the value of the applied magnetic field. In addition, the plasma system becomes stable because the strongly generated magnetic field overlaps the growth rate, as observed by Rubab *et al.*[42] Due to its frequent occurrence in both experimental as well as astrophysical plasmas, the thermal Weibel instability caused by temperature anisotropy has garnered considerable interest. Researchers have recently shown a great deal of interest in the WI due to its potential significance in scientific and astrophysical plasmas. In galactic plasmas,[30] the Weibel magnetic fields are assumed to be the beginning of turbulence and dynamo amplification. Weibel

26 February 2024 09:20:26

© Author(s) 2024

instability is also believed to play a significant role in a variety of scenarios that involve matter under extremely difficult circumstances, including gamma-ray bursts,[31,32] relativistic jets in active galactic nuclei,[33] collisionless shocks,[34,35] inertial confinement fusion,[36,37] electron-positron,[38,39] and quark–gluon plasmas.[40]

IV. CONCLUSION

In this paper, it is observed that the disturbance of a wave leads to the generation of a Weibel instability as a result of temperature anisotropy in electrons and ions. The growth rate vs applied external magnetic field is plotted for cases involving only electrons and in the presence of both warm and cold ions. The results indicate that the presence of warm and cold ions within the plasma influences the growth rate of instability. Moreover, the growth rate of the instability is attenuated in the presence of an external magnetic field aligned with the flow. The presence of warm ions can exacerbate the instability arising from electron temperature anisotropy, leading to a higher growth rate in the case of heated ions. The study of instability reveals that during the linear growth phase of the magnetic field, there is an exponential increase in its magnitude. Our theoretical work holds diverse applications in laboratory plasmas, astrophysical environments, laser-produced plasmas, and energy transportation, among other fields. The observations of our research can be applied to study the generation of the magnetic field in various fields, including supernovas, space, and galactic cosmic rays. The findings from this study may be particularly relevant for astrophysical systems, such as pulsar winds, where the understanding of the Weibel instability can contribute to a deeper comprehension of magnetic field dynamics.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Twinkle Pahuja: Writing – review & editing (equal). **Amit Kumar**: Resources (equal); Software (equal). **Himank Sagar**: Data curation (equal); Formal analysis (equal). **Ruby Gupta**: Funding acquisition (equal); Investigation (equal); Methodology (equal). **Jyotsna Sharma**: Investigation (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] E. S. Weibel, Phys. Rev. Lett. **2**, 83 (1959).
[2] M. L. Bernet, F. Miniati, S. J. Lilly, P. P. Kronberg, and M. Dessauges-Zavadsky, Nature **454**, 302–304 (2008).
[3] M. V. Medvedev, L. O. Silva, M. Fiore, R. A. Fonseca, and W. B. Mori, J. Korean Astron. Soc. **37**, 533–541 (2004).
[4] M. V. Medvedev and A. Loeb, Astrophys. J. **526**, 697–706 (1999).
[5] A. Gruzinov and E. Waxman, Astrophys. J. **511**, 852–861 (1999).
[6] C. M. Huntington, F. Fiuza, J. S. Ross, A. B. Zylstra, R. P. Drake, D. H. Froula, G. Gregori, N. L. Kugland, C. C. Kuranz, M. C. Levy, C. K. Li, J. Meinecke, T. Morita, R. Petrasso, C. Plechaty, B. A. Remington, D. D. Ryutov, Y. Sakawa, A. Spitkovsky, H. Takabe, and H.-S. Park, Nat. Phys. **11**, 173–176 (2015).
[7] N. G. Zaki, Pramana **75**, 719–726 (2010).
[8] A. Achterberg and J. Wiersma, Astron. Astrophys. **475**(1), 1–18 (2007).
[9] M. Lazar, R. Schlickeiser, S. Poedts, and R. C. Tautz, Mon. Not. R. Astron. Soc. **390**, 168–174 (2008).
[10] L. V. Borodachev and D. O. Kolomiets, J. Plasma Phys. **77**, 277–287 (2010).
[11] D. D. Ryutov, F. Fiuza, C. M. Huntington, J. S. Ross, and H.-S. Park, Phys. Plasmas **21**, 032701 (2014).
[12] M. Mahdavi and H. Khanzadeh, Adv. High Energy Phys. **2015**, 428013.
[13] Aman-ur-Rehman, S. Ali Shan, and T. Majeed, Phys. Plasmas **24**, 122113 (2017).
[14] A. Grassi, M. Grech, F. Amiranoff, F. Pegoraro, A. Macchi, and C. Riconda, Phys. Rev. E **95**, 023203 (2017).
[15] T. Silva, B. Afeyan, and L. O. Silva, Phys. Rev. E **104**, 035201 (2021).
[16] C. Zhang, Y. Wu, M. Sinclair, A. Farrell, K. A. Marsh, J. Hua, I. Petrushina, N. Vafaei-Najafabadi, R. Kupfer, K. Kusche, M. Fedurin, I. Pogorelsky, M. Polyanskiy, C.-K. Huang, W. Lu, W. B. Mori, and C. Joshi, Phys. Plasmas **29**, 062102 (2022).
[17] C. Zhang, Y. Wu, M. Sinclair, A. Farrell, K. A. Marsh, I. Petrushina, N. Vafaei-Najafabadi, A. Gaikwad, R. Kupfer, K. Kusche, M. Fedurin, I. Pogorelsky, M. Polyanskiy, C. K. Huang, J. Hua, W. Lu, W. B. Mori, and C. Joshi, Proc. Natl. Acad. Sci. U. S. A. **119**, 50 (2022).
[18] A. Kumar, R. Gupta, and J. Sharma, AIP Adv. **12**, 065013 (2022).
[19] A. Kumar, M. Yadav, R. Gupta, and J. Sharma, AIP Conf. Proc. **2745**, 030004 (2023).
[20] M. Ghorbanalilu, Plasma Phys. Controlled Fusion **55**, 045002 (2013).
[21] C. F. Kennel and R. Z. Sagdeev, J. Geophys. Res. **72**, 3303, https://doi.org/10.1029/jz072i013p03303 (1967).
[22] N. T. Gladd, Phys. Fluids **26**, 974 (1983).
[23] M. S. Wei, F. N. Beg, E. L. Clark, A. E. Dangor, R. G. Evans, A. Gopal, K. W. D. Ledingham, P. McKenna, P. A. Norreys, M. Tatarakis, M. Zepf, and K. Krushelnick, Phys. Rev. E **70**, 056412 (2004).
[24] M. Ghorbanalilu, Plasma Phys. Controlled Fusion **53**, 035006 (2011).
[25] G. A. Mourou, T. Tajima, and S. V. Bulanov, Rev. Mod. Phys. **78**, 309 (2006).
[26] M. Marklund and P. K. Shukla, Rev. Mod. Phys. **78**, 591 (2006).
[27] P. Sharma, K. Avinash, and D. N. Gupta, Laser Part. Beams **36**(3), 376–383 (2018).
[28] D. N. Gupta, H. J. Jang, and H. Suk, J. Appl. Phys. **105**(10), 106110 (2009).
[29] L. O. Silva, R. A. Fonseca, J. W. Tonge, J. M. Dawson, W. B. Mori, and M. V. Medvedev, Astrophys. J. **596**, L121 (2003).
[30] R. M. Kulsrud and S. W. Anderson, Astrophys. J. **396**, 606 (1992).
[31] M. S. Hur, D. N. Gupta, and H. Suk, Phys. Lett. A **372**(15), 2684–2687 (2007).
[32] Y. Lyubarsky and D. Eichler, Astrophys. J. **647**, 1250 (2006).
[33] K.-I. Nishikawa, P. Hardee, G. Richardson, R. Preece, H. Sol, and G. J. Fishman, Astrophys. J. **595**, 555 (2003).
[34] D. Caprioli and A. Spitkovsky, Astrophys. J. **765**, L20 (2013).
[35] R. Blandford and D. Eichler, Phys. Rep. **154**(1), 1–75 (1987).
[36] A. Macchi, A. Antonicci, S. Atzeni, D. Batani, F. Califano, F. Cornolti, J. J. Honrubia, T. V. Lisseikina, F. Pegoraro, and M. Temporal, Nucl. Fusion **43**, 362 (2003).
[37] L. O. Silva, R. A. Fonseca, J. W. Tonge, W. B. Mori, and J. M. Dawson, Phys. Plasmas **9**, 2458 (2002).
[38] C. Ren, M. Tzoufras, F. S. Tsung, W. B. Mori, S. Amorini, R. A. Fonseca, L. O. Silva, J. C. Adam, and A. Heron, Phys. Rev. Lett. **93**, 185004 (2004).
[39] T.-Y. B. Yang, J. Arons, and A. B. Langdon, Phys. Plasmas **1**, 3059 (1994).
[40] R. A. Fonseca, L. O. Silva, J. Tonge, R. G. Hemker, J. M. Dawson, and W. B. Mori, IEEE Trans. Plasma Sci. **30**, 28 (2002).
[41] P. Arnold, J. Lenaghan, G. D. Moore, and L. G. Yaffe, Phys. Rev. Lett. **94**, 072302 (2005).

26 February 2024 09:20:26

© Author(s) 2024

[42]N. Rubab, A. C.-L. Chian, and V. Jatenco-Pereira, J. Geophys. Res.: Space Phys. **121**, 1874–1885, https://doi.org/10.1002/2015ja022206 (2016).
[43]S. C. Sharma, J. Sharma, A. Bhasin, and R. Walia, J. Plasma Phys. **78**, 635–640 (2012).
[44]S. C. Sharma, D. Kaur, A. Gahlot, and J. Sharma, Phys. Plasmas **21**, 103702 (2014).
[45]J. Sharma, S. C. Sharma, V. K. Jain, and A. Gahlot, J. Plasma Phys. **79**, 577–585 (2013).
[46]A. Kumar, R. Gupta, and J. Sharma, AIP Adv. **11**, 115022 (2021).
[47]A. Kumar, R. Gupta, and J. Sharma, AIP Adv. **12**, 035026 (2022).
[48]A. Kumar, R. Dhawan, R. Gupta, and J. Sharma, Mater. Today: Proc. **62**(6), 3263–3268 (2022).
[49]A. Kumar, R. Dhawan, R. Gupta, N. Sabharwal, and J. Sharma, E3S Web Conf. **430**, 01272 (2023).
[50]A. Kumar and J. Sharma, AIP Conf. Proc. **2821**, 060004 (2023).
[51]J. Sharma, S. Singh, A. Kumar, M. A. Hossain, S. H. Ahammad, and A. N. Z. Rashed, Braz. J. Phys. **53**, 128 (2023).
[52]J. Sharma, A. Kumar, R. Gupta, S. Vijayalakshmi, L. M. Leo, V. Kannan, M. A. Hossain, S. H. Ahammad, and A. N. Z. Rashed, Iran. J. Sci. **47**(6), 01559 (2023).

26 February 2024 09:20:26

© Author(s) 2024

Electrical Engineering
https://doi.org/10.1007/s00202-024-02245-7

ORIGINAL PAPER

Investigation and mitigation of unbalanced hall sensor signal faults in sensored brushless DC motor drives

Arvind Goswami[1] · **Mini Sreejeth**[1] · **Madhusudan Singh**[1]

Received: 16 May 2023 / Accepted: 6 January 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
This paper investigates hall sensor signal unbalancing faults in sensored Brushless Direct Current (BLDC) motor drives and presents a fault effect minimization scheme for implementation in motor controller without any modification in system architecture, or core commutation method. In the proposed scheme detection, classification, and estimation of fault extent are carried out separately at the start of controller itself for subsequent corrections without involving issue of repetitive error estimation and finding valid reference signal separately for correcting inverter switching mistiming, as required in classical methods. The categorization part of scheme maps the faults with corresponding hall sensors. The proposed scheme is validated in MATLAB/Simulink model. It is tested on experimental test bench of controller and 1 kW BLDC motor of an electric rickshaw, along with another technique based on least square method (LSM) for performance comparison. Results show reduction of peak-to-peak value by 30% and ripple content by 50% in phase currents, and peak-to-peak reduction of 47% with 29% lesser ripple content in DC link current by using the presented scheme. The observed increment in efficiency is 0.8% and 2% in nominal torque/ampere at 30% load. At 100% load, increase of 2.2% efficiency and 9% nominal torque/ampere are observed. The increased memory loading of controller is estimated to be 1.25 times as compared to normal run without any correction scheme. Approximately 24% of this additional loading is utilized for repetitive calculations as compared to 63% in the other LSM based scheme, resulting in lesser run time loading of controller. The transient response with sudden load change shows better response with implementation of the proposed scheme in controller.

Keywords Brushless DC motor · Fault classification · Fault-tolerant control · Hall sensor fault · Motor controller · Sensored brushless DC motor drive

1 Introduction

Brushless DC (BLDC) motors with hall effect sensors (HESs) are one of the dominantly used arrangements in low to medium power electric vehicles (EV), industrial automation and robotics [1]. These motors have high torque to weight ratio and efficiency with precise discrete rotor position detection and control mechanism [2, 3]. BLDC motors can be operated in sensored mode with three HESs or in sensorless mode without any HES. EV applications like two wheelers, electric rickshaw and tuk-tuk take advantage of the cost-effective drive mechanism using BLDC motor and low cost HESs. The controllers used in these applications utilize HES signals to control switching timings sequence [4, 5]. The drive control mechanism is dependent on integrity of HES signals and their synchronization with Back Electromotive Force (BEMF) profile of motor [6]. However, deviations in HES signal from ideal timings are observed due to factors such as ageing, mounting processes of HESs, windings and placement of the sensors. This is often reported in mass scale produced motors particularly for low voltage EV applications and is reported in experimental studies [7, 8]. Due to limitations and difficulties in assembling stator lamination sheets with HESs precisely in various slot numbering per pole fixtures, net signal positioning of sensors is not precise in many cases. The net effect is even more pronounced in machines with higher number of poles leading to high electrical degree error even for small unbalancing errors [9–11].

Various fault tolerant control (FTC) schemes for sensored control of BLDC motors that utilize phase angle advancement and signal reconstruction, based on the concept of

✉ Arvind Goswami
arvind_phd2k19@dtu.ac.in

[1] Department of Electrical Engineering, Delhi Technological University, Delhi, India

Published online: 12 February 2024

instantaneous speed averaging, extrapolation or other complex methods are reported in literature for HES faults [1, 12–14]. As per authors' best knowledge, these methods can be modified for implementation in sensored BLDC motor applications. However, the process of signal reconstruction with inherent repeated error estimation for 60° electrical (°E) sectors, increases execution time and complexity, particularly if interrupt subroutine scheduling gets affected which can lead to distortions in applied phase voltage signals [15–18].

This paper proposes an FTC method with detection and correction schemes for unbalanced HES signals in sensored BLDC motor drives. Errors are categorized in one, two or three HES faults and mapped with corresponding HESs, The exact information of type of unbalanced HES signals result in easier and faster rectification of consequent mistimed switching signals. The proposed scheme is developed for seamless integration in controller without any change in circuit, control algorithm, state machine model or restriction in application of other FTC schemes in controller. Detection and mitigation of other run time faults remain unaffected and other FTC methods can be used as per control algorithm and state machine model requirements.

The overall contribution of this paper and utility of the proposed scheme are summarized as follows:

- Unbalanced HES signals and their effects on sensored BLDC motor drive operation are investigated.
- FTC scheme with signal analysis and information extraction is proposed for easier implementation and error effect minimization.
- Method to detect, classify and map mistimed HES signals to corresponding HESs is proposed.
- Experimental validation of the proposed scheme in actual controller is presented and it shows superior steady state and dynamic performance over faulty condition and performs better against another method based on LSM.

This paper is structured into seven sections including introduction in Sect. 1. Sections 2 and 3 describe the system operation and unbalanced hall sensor signal faults with proposed categorization scheme. Detection and correction scheme application in controller is elaborated in Sect. 4, followed by simulation results in Sect. 5. Section 6 presents the experimental results and discussion. Final conclusions are drawn in Sect. 7.

2 System operation

HES output voltage signals are dependent on applied magnetic field strength which varies according to the rotation of rotor magnetic poles [19]. Standard 3 HES configuration provides six valid combinations of binary output signals in an electrical cycle, dividing it into six sectors of 60°E span. Rotor position and rotation direction are tracked by changes in this signal pattern [20]. Inverter switching sequence is dependent on sector number and switching signals that are synchronized with HES transition signals [21].

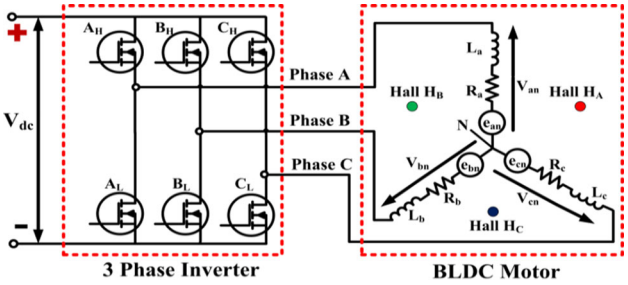


Fig. 1 BLDC motor and inverter circuit with hall sensors

Figure 1 shows the inverter and switching scheme referred in this paper. Corresponding 120° trapezoidal commutation switching sequence with HES signal pattern is presented in Table 1. Figure 2 shows phase voltage and phase current pattern against HES signals with BEMF pattern for 120° trapezoidal commutation. The 60°E sector are denoted by S_i ($i = 1$ to 6). The major functions in a sensored BLDC drive affected by unbalanced HES signals are:

- Calculation of rotor speed and PWM duty cycle.
- Inverter switching pattern determination according to commutation sequence.
- Detection of forward and reverse rotor rotation.
- Motor starting sequence in controller.

3 Unbalanced hall sensor signal faults and proposed categorization

The HESs in ideal case can be represented by three equidistant sensors to represent exact 60°E sectors. The cases with unbalanced HES signals are shown by moving each HES by angular error equivalent degree from ideal placement. The effect on 60°E sectors is shown in Figs. 3, 4, 5 and 6 for different cases. Ideal hall placements (H_A, H_B, H_C) are shown with solid lines and, dotted lines represent HESs with error. Angular deviations are represented as Φ_a, Φ_b and Φ_c, respectively. HES mistiming faults are further categorized in the following cases for easier detection and rectification.

3.1 Case 1: Single hall signal unbalancing

In this case, one HES signal is mistimed and the other two HES signals are correctly timed. This condition is depicted in Fig. 3 with H_a being the HES with delayed signal. This

Table 1 Switching sequence with hall signal pattern

Sector no	Rotor angle (°E)	Hall sensor signal			Switching pattern						Phase voltage		
i	Φ	H_C	H_B	H_A	A_H	B_H	C_H	A_L	B_L	C_L	Phase C	Phase B	Phase A
1	0°–60°	0	1	0	1	0	0	0	1	0	0	$-V_{dc}$	$+V_{dc}$
2	60°–120°	1	1	0	1	0	0	0	0	1	$-V_{dc}$	0	$+V_{dc}$
3	120°–180°	1	0	0	0	1	0	0	0	1	$-V_{dc}$	$+V_{dc}$	0
4	180°–240°	1	0	1	0	1	0	1	0	0	0	$+V_{dc}$	$-V_{dc}$
5	240°–300°	0	0	1	0	0	1	1	0	0	$+V_{dc}$	0	$-V_{dc}$
6	300°–360°	0	1	1	0	0	1	0	1	0	$+V_{dc}$	$-V_{dc}$	0

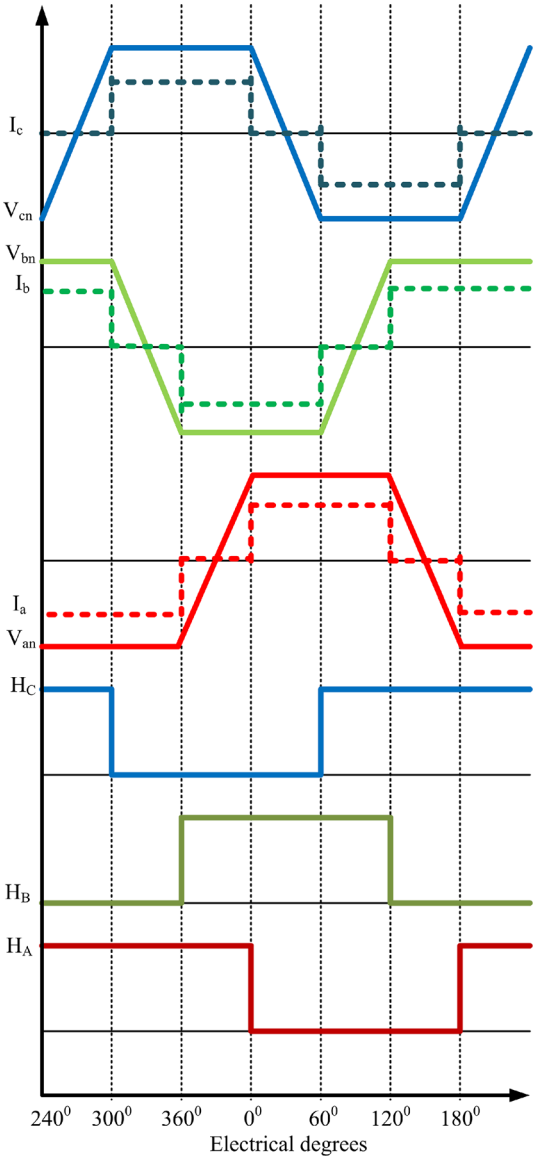


Fig. 2 Phase voltage, stator current and BEMF profiles with HES transition signals referred to rotor rotation angle

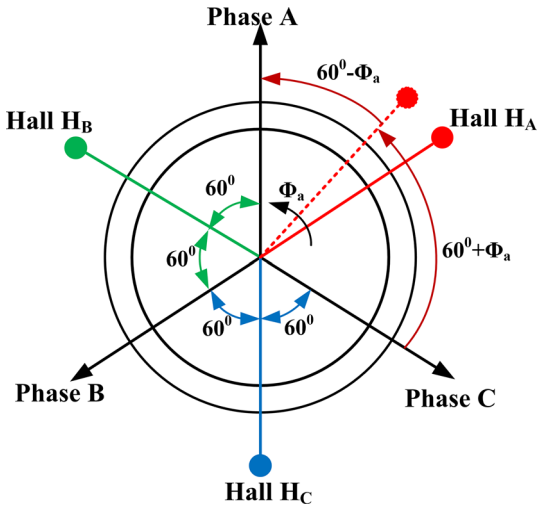


Fig. 3 Cross sectional representation of equivalent HES unbalancing due to single hall mistimed signal

condition results in one electrical sector extending to $60° + \Phi_a$ and the adjacent sector shrinking to $60° - \Phi_a$ where Φ_a is the equivalent HES signal unbalance. Thus, two sector spans different from other four sector spans with opposite and same degree represent the case of single HES signal mistiming.

The observed cyclic sector span deviation pattern for both directions of rotation in Case 1 can be represented as:

$$+\Phi_a, -\Phi_a, n, n, n, n$$
$$or -\Phi_a, +\Phi_a, n, n, n, n$$

where " + " and "−" signs represent sector span elongation and compression, respectively, and "n" normal sector span.

3.2 Case 2(A): Two HES signals unbalanced in same direction

The condition with one correctly timed and two mistimed HES signals results in alteration in four consecutive sector spans. The rest two consecutive sectors retain their normal

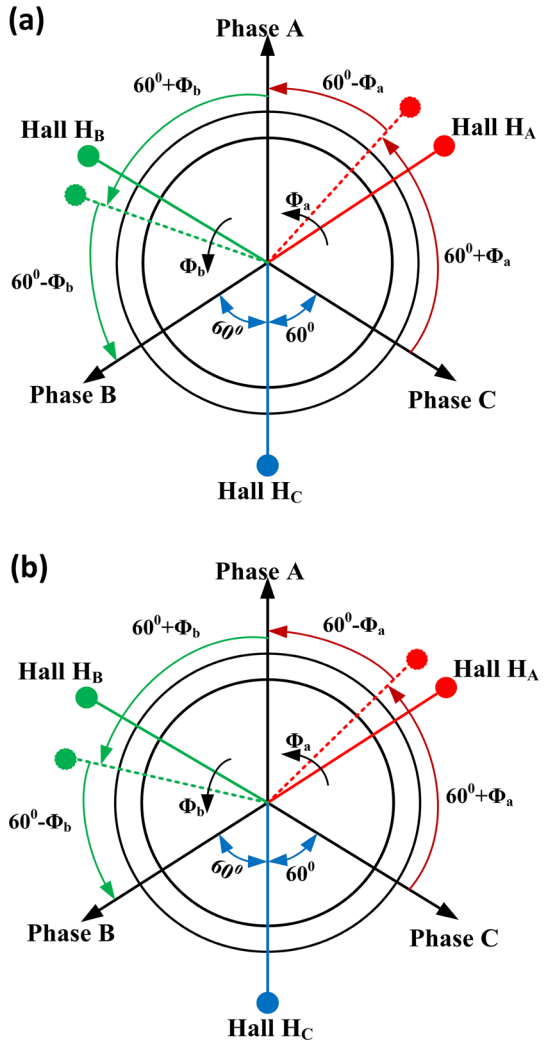


Fig. 4 Equivalent misalignment due to two unbalanced HES signals in same direction: **a** $\Phi_a > \Phi_b$; **b** $\Phi_a < \Phi_b$

span. This results in two conditions shown in Fig. 4 wherein Φ_a is the 1st deviation while traversing from normal span sectors to altered span sectors, and Φ_b is the 2nd deviation. Figure 4a shows the condition when $\Phi_a > \Phi_b$. Figure 4b shows the scenario when $\Phi_a < \Phi_b$. Both these conditions result in $60° + \Phi_a, 60° - \Phi_a$ and $60° + \Phi_b, 60° - \Phi_b$ sector spans. Thus, sector span deviation pattern in Case 2(A) with unbalanced HESs H_A and H_B can be represented as follows:

$$+\Phi_a, -\Phi_a, +\Phi_b, -\Phi_b, n, n$$

In case of reverse direction of rotation, sector span deviation pattern is:

$$-\Phi_b, +\Phi_b, -\Phi_a, +\Phi_a, n, n$$

3.3 Case 2(B): Two HES signals unbalanced in opposite direction

Figure 5 shows two possible conditions when misalignment of both HESs is in opposite direction. This results in sector pattern of $60° - \Phi_a$, $60° + \Phi_a$ and $60° + \Phi_b$, $60° - \Phi_b$ considering counterclockwise direction of rotation. Thus, sectoral span deviation pattern for both possible conditions of Case 2(B) with unbalanced HESs H_A and H_B can be represented as:

$$-\Phi_a, +\Phi_a, +\Phi_b, -\Phi_b, n, n$$
$$\text{or} + \Phi_a, -\Phi_a, -\Phi_b, +\Phi_b, n, n$$

3.4 Case 3: Three HES signal unbalancing

When all three HES signals show unbalancing, resultant sector duration pattern has six sectors as three pairs with deviations as shown in Fig. 6. Each pair has one sector expanded and next sector compressed by same degree. The sector spans in pairs are $60° + \Phi_a$, $60° - \Phi_a$; $60° + \Phi_b$, $60° - \Phi_b$ and $60° + \Phi_c$, $60° - \Phi_c$. Thus, resultant sector span deviation pattern for Case 3 can be represented as follows:

$$+\Phi_a, -\Phi_a; +\Phi_b, -\Phi_b; +\Phi_c, -\Phi_c$$

3.5 Case 4: Three HES signal unbalancing with common error

The particular case when all three HES signals show mistiming but observed deviation is same for all three signals, can be detected and corrected separately. This can happen due to deviation in HES mounting in one direction, moving each HES by certain degree from ideal placement, or run time effects common to all three HESs. The error, even in sub degree level introduced during stator stamping cutting for HES adjustment can result in significant deviation in 60°E sector span. Resultant deviation for Case 4 referred to Fig. 6 with Φ_0 as equivalent HES signal deviation, can be represented as follows:

$$\Phi_a = \Phi_b = \Phi_c = \Phi_0$$
$$+\Phi_0, -\Phi_0; +\Phi_0, -\Phi_0; +\Phi_0, -\Phi_0$$
$$or - \Phi_0, +\Phi_0; -\Phi_0, +\Phi_0; -\Phi_0, +\Phi_0$$

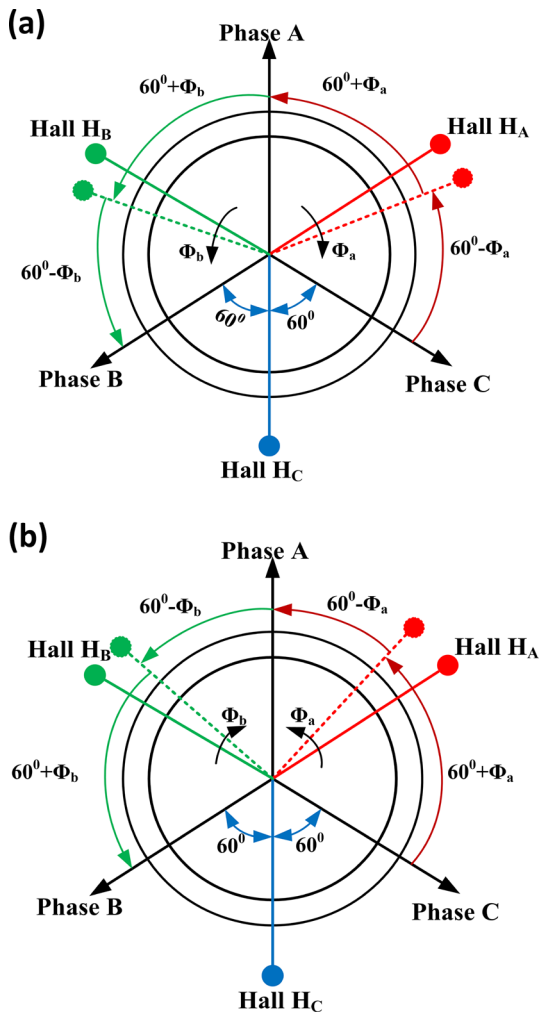


Fig. 5 Equivalent HES unbalancing in opposite directions due to two HESs: **a** divergent case; **b** convergent case

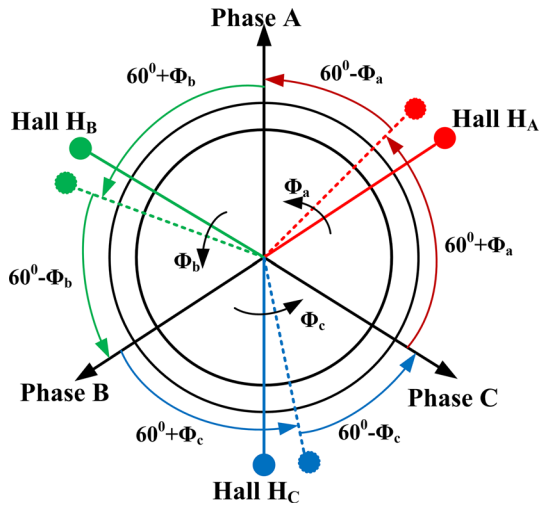


Fig. 6 Sector span pattern with three unbalanced HESs

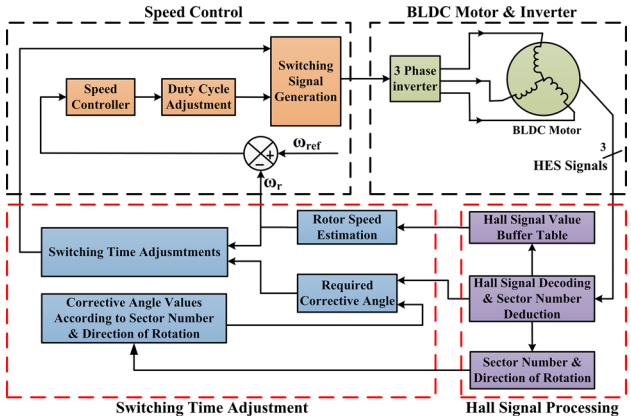


Fig. 7 Motor controller with proposed FTC scheme

4 Detection and correction scheme application in controller

The correction method based on the proposed detection and categorization scheme is developed for implementation in controller during first time motor controller integration. Subsequent operations can implement the estimated corrections without disturbing the state machine model due to computational overheads in microcontroller [21, 22]. The modified controller with inclusion of proposed scheme is presented in Fig. 7.

Figure 8 flowchart explains the application of the proposed scheme in controller. The sub-process of sequential detection of case of HES unbalancing error, highlighted in main flowchart, is further presented in detail in Fig. 9. The presented scheme confirms to the following features for integration in a sensored BLDC motor controller:

- It does not introduce large number of complex new parameters, variables, and lengthy calculations. This is because at higher speeds even small latency results in high amount of switching angle change that may directly disturb the switching logic timings.
- The interrupt subroutine execution time corresponding to switching sequence implementation remains unaffected.
- It does not interfere with other periodic and critical functions of PWM generation, duty cycle estimation, ADC and DMA buffering operations.
- The state machine logic and sequential operations in target application remain unchanged without any new physical or pseudo component in the controller.

The FTC data values are stored in microcontroller's non-volatile memory during motor controller integration process, along with a flag to indicate the applicability of the presented FTC for further operations. Thus, repetition of error detection and estimation is avoided. Once ascertained, the required

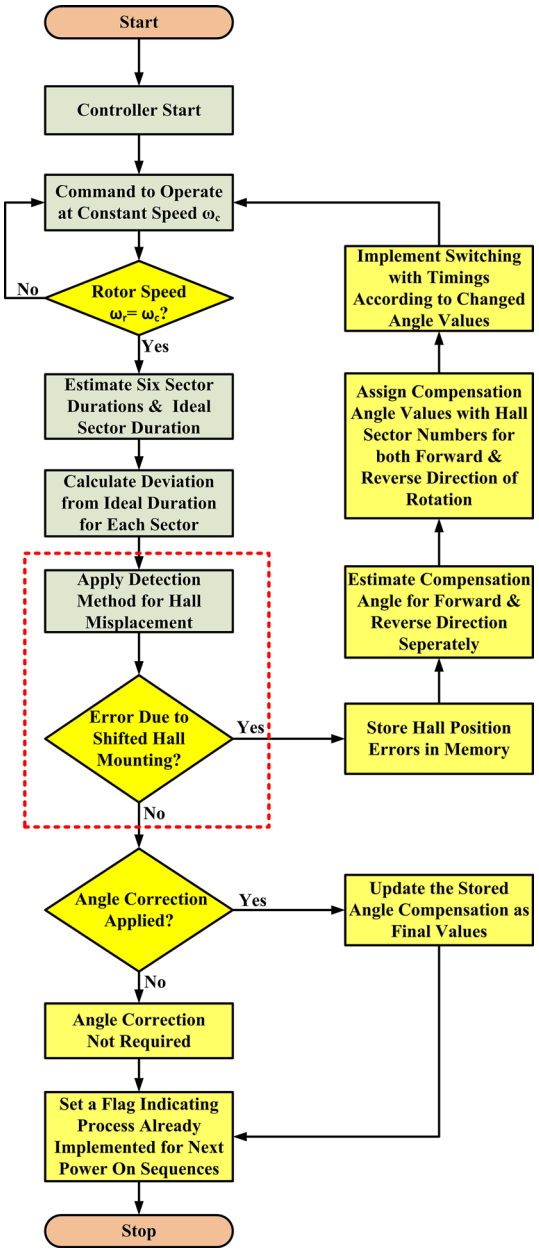


Fig. 8 Flowchart of FTC scheme integration in the motor controller

corrections related to present and upcoming HES signal pattern are stored separately for use in both forward and reverse direction of rotation. The proposed FTC scheme executes following sequential steps in motor controller:

- Sector spans S_i ($i = 1$ to 6) are calculated as follows:

$$S_i = t_{i+1} - t_i. \tag{1}$$

- One electrical cycle time span S_e is calculated as:

$$S_e = \sum S_i = S_1 + S_2 + S_3 + S_4 + S_5 + S_6. \tag{2}$$

- Rotor electrical speed ω_e and mechanical speed ω_r are calculated using Eq. (3) where P is number of pole pairs.

$$\begin{aligned} \omega_e &= (1/S_e) * 60 \\ \omega_{\mathrm{r}} &= (1/P) * (1/S_e) * 60. \end{aligned} \tag{3}$$

- Ideal 60° E sector duration (S_0) is obtained as:

$$S_0 = S_e/6. \tag{4}$$

- Deviations of electrical 60° sectors from ideal sector duration are calculated as follows:

$$\Delta S_i = S_0 - S_i. \tag{5}$$

- Angular sector spans Φ_i are calculated in Eq. (6) and confirm to mutual relation of Eq. (7):
Ideal sector span $= \Phi_0 = 60°$
Angular sector span $= \Phi_i = 60*(S_i/S_0)$

$$\Phi_i = 360 * (S_i/S_e) \tag{6}$$

$$\Phi_1 + \Phi_2 + \Phi_3 + \Phi_4 + \Phi_5 + \Phi_6 = 6\Phi_0 = 360°. \tag{7}$$

- Deviation of each sector from ideal span is calculated using Eq. (8):

$$\Delta\Phi_i = \Phi_0 - \Phi_{\mathrm{i}}. \tag{8}$$

- Cumulative sum of all six sector angular spans equals to one electrical cycle which equals to $6\Phi_0$. Using Eq. (7), the cumulative sum of six consecutive sector deviations is zero as in Eq. (9).

$$\begin{aligned} \sum \Delta\Phi_i &= \Delta\Phi_1 + \Delta\Phi_2 + \Delta\Phi_3 + \Delta\Phi_4 + \Delta\Phi_5 + \Delta\Phi_6 \\ \sum \Delta\Phi_i &= 6\Phi_0 - (\Phi_1 + \Phi_2 + \Phi_3 + \Phi_4 + \Phi_5 + \Phi_6) \\ \sum \Delta\Phi_i &= 0. \end{aligned} \tag{9}$$

- Detection and categorization of fault are based on the cases presented in Figs. 3, 4, 5 and 6 case. The condition for Case 1 fault in Fig. 4 is represented in Eqs. (11, 12).

$$\Delta\Phi_i = -\Delta\Phi_{i+1} \tag{10}$$

$$\Delta\Phi_1 = \Delta\Phi_3 = \Delta\Phi_5 = \Delta\Phi_6 = 0 \tag{11}$$

$$\Delta\Phi_2 = -\Delta\Phi_3 \tag{12}$$

- In Case 2(A) two starting sectors with deviations must be of opposite signs and same magnitude and next two sector deviations separately must be of opposite sign with same magnitudes, represented in Eq. (13) subject to relation in Eq. (11).

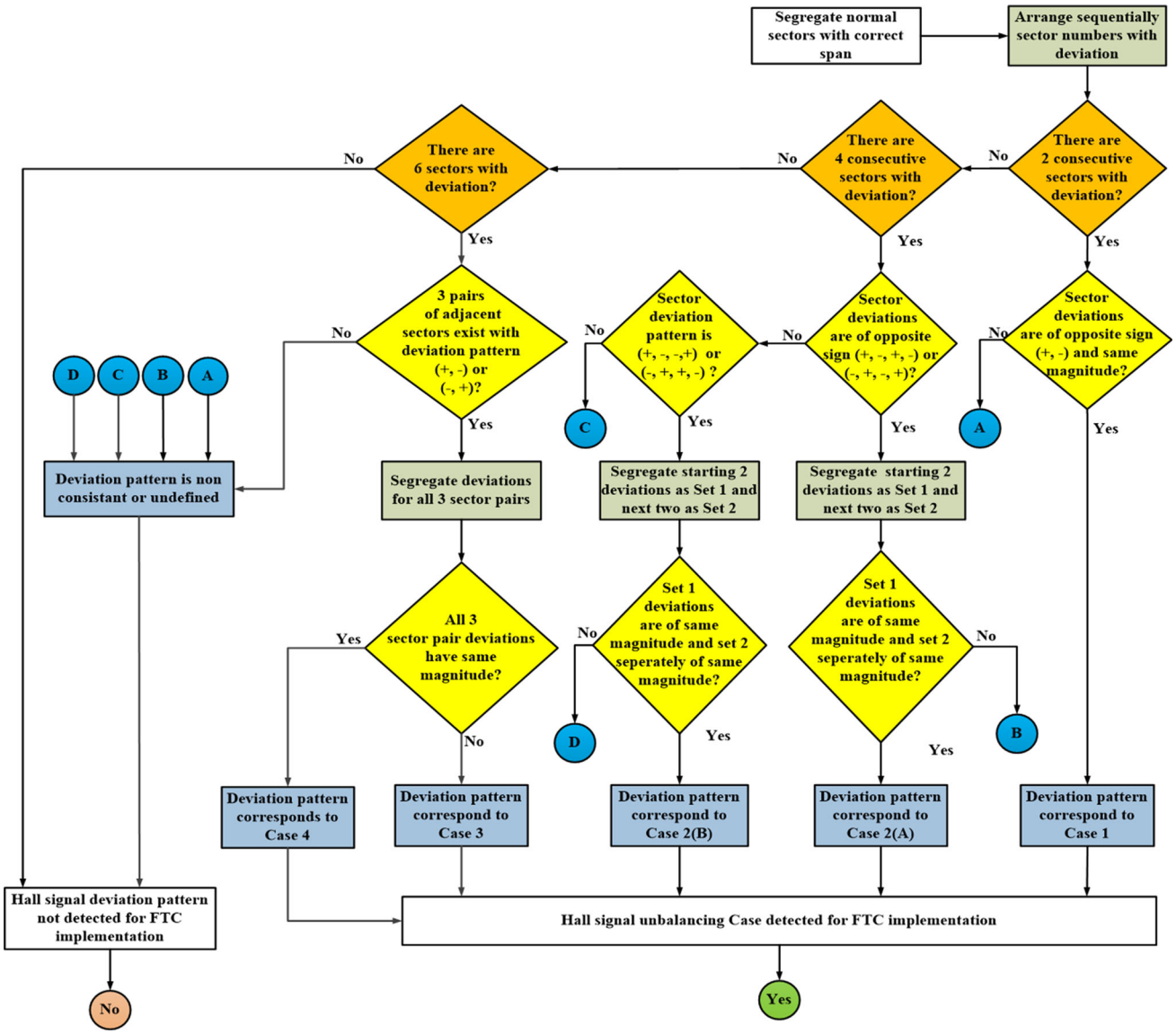


Fig. 9 Detection and categorization sub-process flowchart for unbalanced HES signals for integration in motor controller

$$\Delta\Phi_i = -\Delta\Phi_{i+1};\ \Delta\Phi_{i+2} = -\Delta\Phi_{i+3}$$
$$\Delta\Phi_{i+4} = \Delta\Phi_{i+5} = 0 \tag{13}$$

The observed deviation pattern for Fig. 5 is:

$$\Delta\Phi_2 = -\Delta\Phi_3 \text{ and } \Delta\Phi_4 = -\Delta\Phi_5$$
$$\Delta\Phi_1 = \Delta\Phi_6 = 0 \tag{14}$$

- The Case 2(B) condition is represented in Eq. (15) subject to relation in Eq. (11):

$$\Delta\Phi_i = -\Delta\Phi_{i+1};\ \Delta\Phi_{i+2} = -\Delta\Phi_{i+3}$$
$$\Delta\Phi_{i+1} * \Delta\Phi_{i+2} > 0 \tag{15}$$
$$\Delta\Phi_{i+4} = \Delta\Phi_{i+5} = 0$$

- In Case 3 the deviations in consecutive sectors in a pair must be of opposite sign and same degree, as in Eq. (16) subject to relation in Eq. (11):

$$\Delta\Phi_i = -\Delta\Phi_{i+1};\ \Delta\Phi_{i+2} = -\Delta\Phi_{i+3},\ \Delta\Phi_{i+4} = -\Delta\Phi_{i+5}. \tag{16}$$

- The condition for Case 4 where all three HES signals exhibit equal unbalancing is expressed in Eq. (17) as follows:

$$\Delta\Phi_i = -\Delta\Phi_{i+1} = \Delta\Phi_{i+2} = -\Delta\Phi_{i+3} = \Delta\Phi_{i+4} = -\Delta\Phi_{i+5} \tag{17}$$

Table 2 Detection and categorization pattern check

Implemented				Detected			
Case	Error (°E)			Case	Error (°E)		
1	0	0	10	1	0	0	10.37
2(A)	0	4	10	2(A)	0	4.02	10.04
2(B)	0	− 6	10	2(B)	0.01	− 5.98	10.01
3	− 6	10	6	3	− 6.06	9.98	6.03
4	20	20	20	4	20.02	19.96	20.01

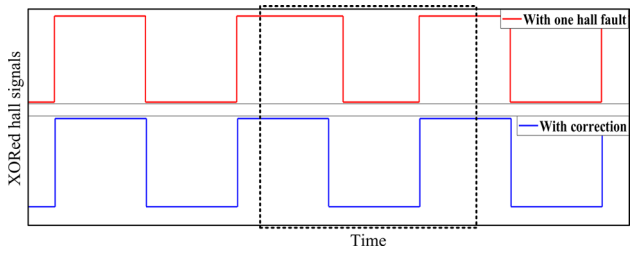


Fig. 10 Case 1 sector extension and compression pattern

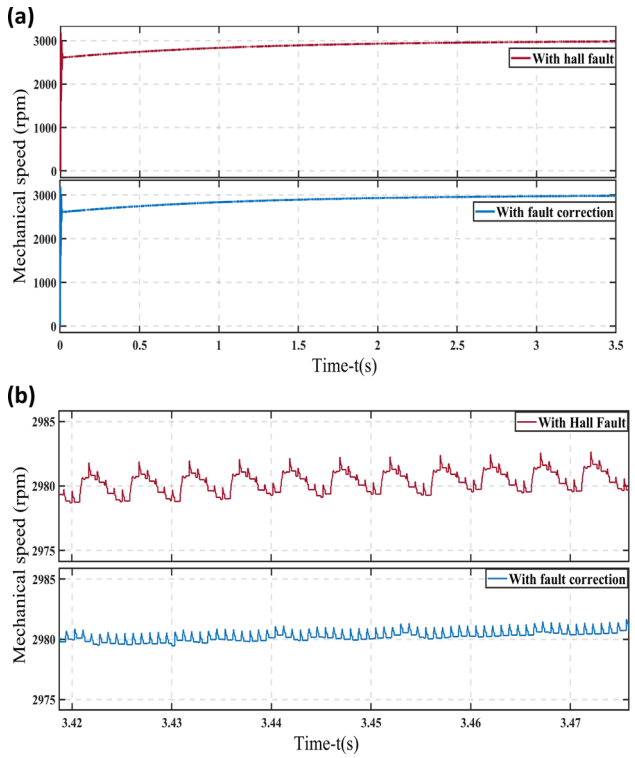


Fig. 11 Rotor speed profile at step load: **a** full range profile; **b** enlarged speed variations

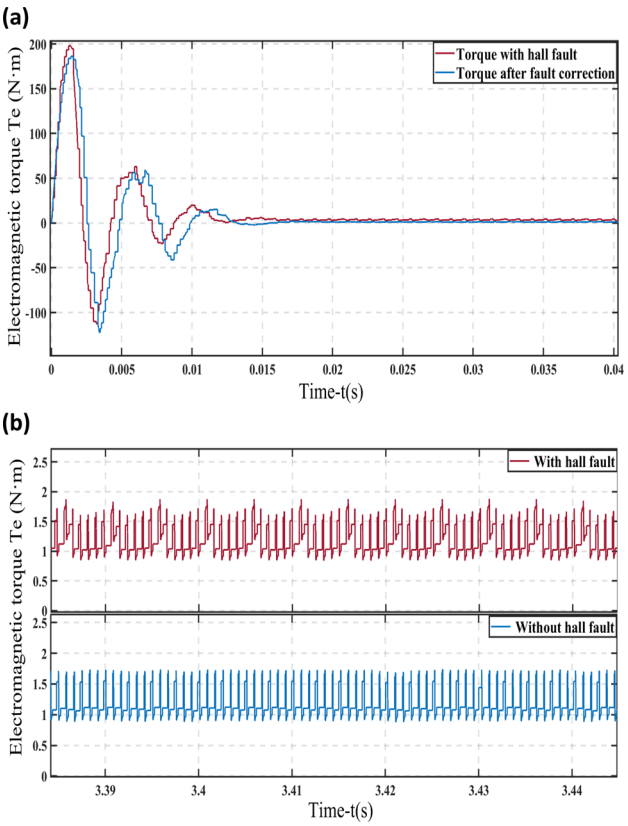


Fig. 12 Electromagnetic torque at step load: **a** full range profile; **b** torque ripple profile

5 Simulation results

The proposed FTC scheme is modeled and simulated in MATLAB software on a 1 kW, 3 phase, 4 pole pair BLDC motor with 3000 RPM speed. Summary of equivalent error detection and category check from HES signals is presented in Table 2 and shown in Fig. 10. The case described in Sect. 3 is correctly identified and corrected accordingly. HES signal pattern of Fig. 10 confirms Case 1. The difference between both cases when corresponding correction is implemented at unit step load is shown in Figs. 11, 12 and 13. The correction in speed profile is shown in Fig. 11. Torque ripples reduction is shown in Fig. 12b. Corresponding phase currents are shown in Fig. 13. Peak-to-peak current variations also show reduction in Fig. 14 with application of the proposed FTC scheme.

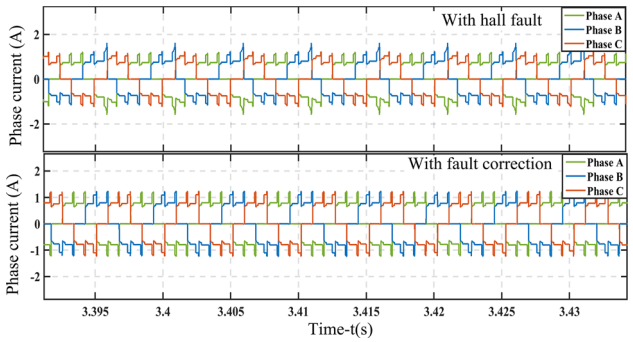


Fig. 13 Three phase stator current profile at step load

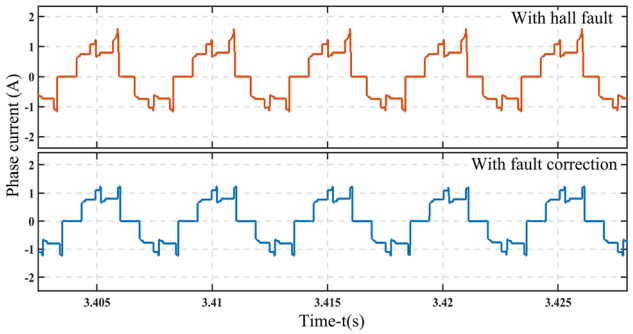


Fig. 14 Single phase stator current distortions at step load

Table 3 Experimental setup details

Rated motor parameters		Test bench setup	
Power	1 kW	DC supply	48 V,72 Ah
Voltage	48 V	Controller	STM32301
Speed (RPM)	3000	Switching method	120° Trapezoidal
Pole pairs	4	Hall sensors	3
Winding	3 Phase star	PWM frequency	16 kHz

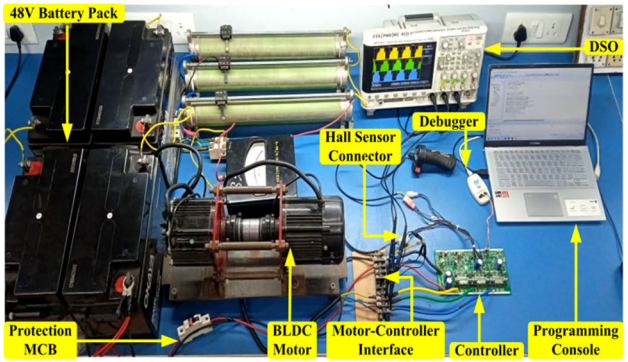


Fig. 15 BLDC motor controller experimental set up

6 Experimental results and discussion

The proposed scheme is tested using 1 kW BLDC motor used in electric autorickshaw. The motor controller is developed around 32-bit STM32 ARM microcontroller. Inverter switching scheme is based on 120° Trapezoidal commutation. Experimental test bench setup is shown in Fig. 15 and its configuration is in Table 3. The HES signal error is investigated for detection and identification. Motor controller response is compared in both conditions, with and without FTC implementation.

The fault pattern detected according to the proposed scheme corresponds to Case 4. The observed effects of this HES signal unbalancing on sectors and switching timing pattern are shown in Fig. 16. Sector boundaries are denoted as t_i ($i = 1$ to 6) and high voltage side switching signals observed at microcontroller PWM output port are shown for each phase. The switching pattern of Fig. 16 leads to non-ideal voltage profiles at motor phase terminals when measured against DC link −ve terminal as reference. Elongation of one sector and compression of next one is observed. BEMF is abruptly cut before reaching zero crossing point and appears as sudden voltage in other phase, creating non-ideal situation for phase current, phase voltage and torque. The angular shifts detected for each sector in electrical degrees are presented and summarized in Table 4. The net value of angular shift observed Φ_0 is 21.37°E.

The measured motor phase voltage profiles with unbalanced HES signals without any correction are presented in Fig. 17a and, with correction in Fig. 17b. The enlarged view of phase voltage waveforms for one complete positive conduction cycle, including rising and falling of BEMFs up to zero crossing points are also shown for comparison. The average phase voltage for Fig. 17a is 20.29 V which increases to 21.23 V by applying the proposed scheme, showing a net increase of approximately 4.6%. Similarly, DC RMS component also increases by approximately 3% from 28.45 V to 29.3 V using proposed FTC scheme.

Abrupt phase voltage switching off and on cause distortions in phase currents as shown in Fig. 18a. From the FTC improved current profile in Fig. 18b, it is observed that phase current shows nearly 30% reduction in peak-to-peak values, estimated AC RMS component reduction by approximately 46% and ripple content reduction by 50%, respectively, using proposed FTC method. Peak values of phase current are higher compared to net current drawn from DC battery source because peak values correspond to sudden short duration current rise and fall over the nominal current profile envelope. Since the same hardware set up and inverter is used for all study cases, these current profiles are used for performance comparison.

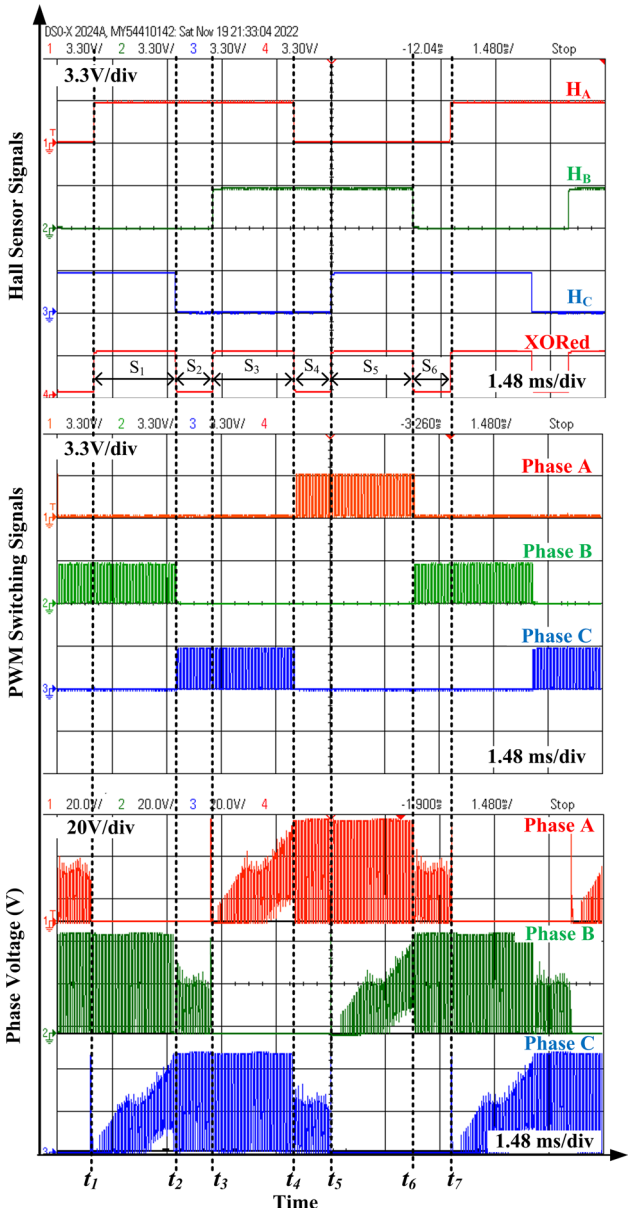


Fig. 16 Motor phase voltage profile with PWM switching signals according to unbalanced HES signals and sectors

DC link current profile is shown in Fig. 19a and b wherein approximately 47% reduction in peak to peak and 13% reduction in average current is observed using FTC method. The AC RMS is reduced by 38% and ripple content is down by nearly 29%.

The validity of the proposed scheme is further checked, and performance compared with another existing method. Method 1 refers to the proposed scheme implementation, Method 2 operation without any correction and Method 3 pertains to the existing scheme implementation for FTC. Method 3 uses First-order Taylor (FOT) series to estimate the electrical angle and linear interpolation based on the least square method (LSM) to estimate average speed with initial angle for the next electrical period [10, 11]. This leads to angular correction to be re-calculated for each electrical cycle with or without changes in speed.

The performance comparison is summarized in Table 5 for implementation of proposed scheme and existing method. It is observed that the proposed method has the least ripple content in both current and voltage along with least DC link current. Peak to peak DC link and phase currents are also minimum with the proposed scheme. The average phase voltage observed for Method 1 is 21.23 V; while, it is 20.29 V and 15.61 V, respectively, for Method 2 and Method 3. The corresponding measured DC RMS voltage is 29.3 V, 28.5 V and 23.44 V for Method 1, Method 2 and Method 3, respectively. This is due to adjustments in durations of 120° conduction periods in Method 1 as shown in Fig. 17. Further, abrupt ending of BEMF rising and falling sections is also reduced in Method 1. Thus, improving the average and DC component in the phase voltage.

Further implementation in STM32 Microcontroller from the memory loading perspective is compared in Table 6. Memory usage and loading values are obtained from IAR software development platform for STM32 ARM controller. Additional memory required using proposed Method 1 is approximately 25% more as compared to Method 2. However, only 24% of it is used for continuously repetitive calculations in controller and most of the calculations are handled during start of controller, which leads to single execution controller load part to nearly 76%. Method 3 has a similar additional memory requirement of 26% but it requires more repetitive computational effort as compared to the proposed method.

Further performance comparison is carried out at 30% and 100% loading of the motor and summarized in Table 7 and Table 8, respectively. DC supply current ripple content is minimum in all conditions with proposed method. It is less by 20% at 30% load and by 12% at full load. This is on the expected line for BLDC motor at full load during stable

Table 4 Check table for observed angular deviation pattern

Sector (S_i)	S_1	S_2	S_3	S_4	S_5	S_6
$\Delta\Phi_i$, (°E)	− 21	+ 21.34	− 21	+ 21.3	− 22.28	+ 21.3
$\Delta\Phi_i$, $i = 1,3,5$	− ve		− ve		− ve	
$\Delta\Phi_i$, $i = 2,4,6$		+ ve		+ ve		+ ve

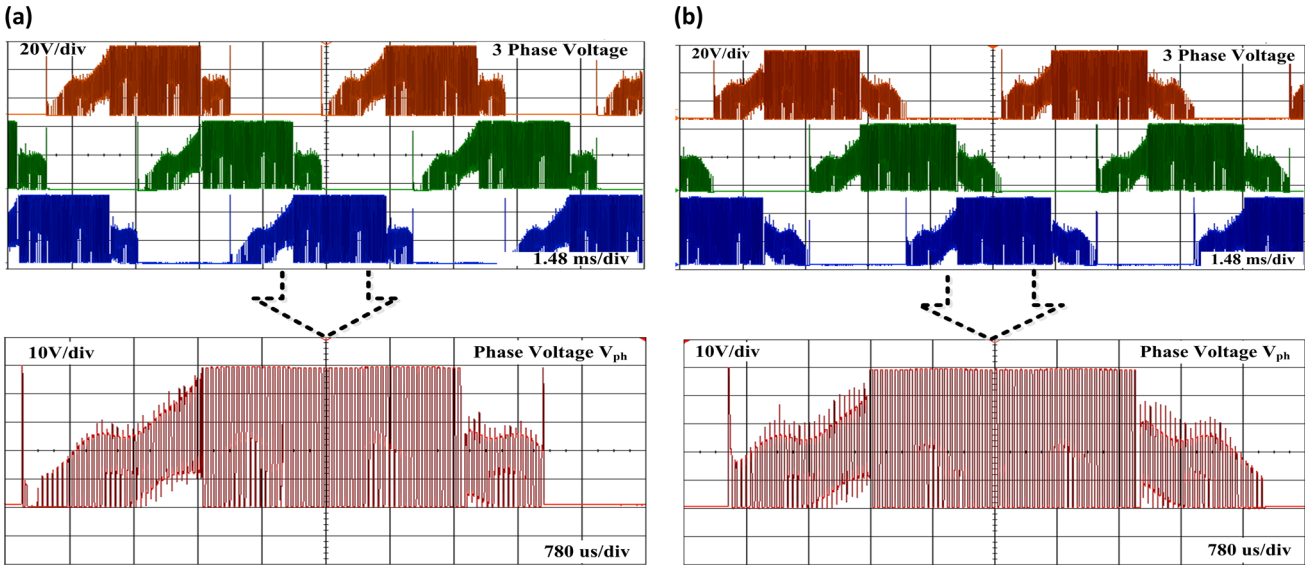


Fig. 17 Measured voltage profile of BLDC motor with unbalanced HES signal: **a** without correction; **b** with correction

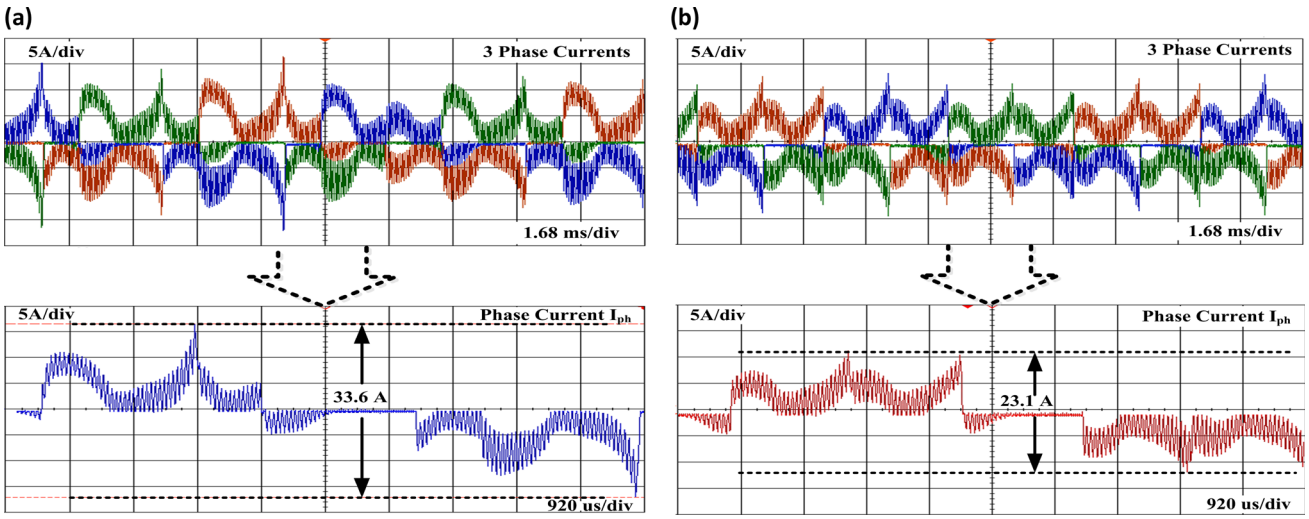


Fig. 18 Measured three phase and single-phase current profile of BLDC motor with unbalanced HES signals: **a** without correction; **b** with correction scheme implementation

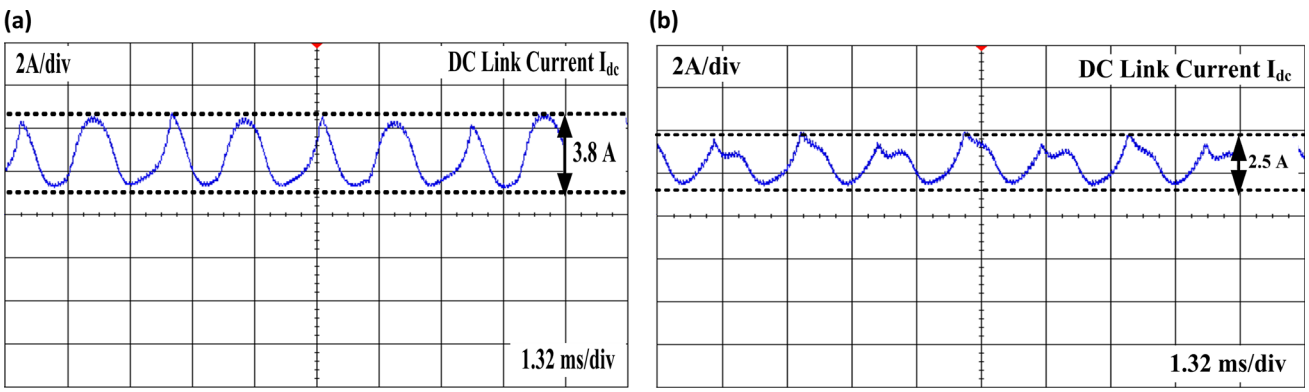


Fig. 19 Measured DC link current with peak-to-peak variation: **a** without correction; **b** with correction scheme implementation

Table 5 Controller implementation comparison

	Phase current (A)			DC link current (A)			Phase voltage (V)		
	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3
Peak to peak	23.1	32.8	42.27	2.5	3.8	4.10	49	49	49
Average	4.150	16.5	0.89	2.401	2.655	2.64	21.23	20.29	15.61
DC RMS	4.8	4.47	5.38	2.5	2.87	2.59	29.3	28.5	23.44
AC RMS	2.398	4.46	5.28	0.667	1.079	1.06	20.3	19.99	17.75
Ripples	49.95%	98.9%	98%	26.68%	37.59%	40.88%	69.28%	70.14%	75.70%

Table 6 Memory utilization comparison

Method	Memory (bytes)	Extra memory required	Repetitive execution part (%)	Single execution part (%)
1	23,601	4,676	24	76
2	18,925	0	0	0
3	23,897	4,972	64.3	35.7

Table 7 Experimental result comparison at 30% load

	Phase current (A)			DC link current (A)			Phase voltage (V)		
	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3
Peak to peak	46	47	64	7.6	10.9	9.6	51	51	51
Average	0.5	0.02	1.76	5.72	5.82	5.96	23.80	23.63	23.17
DC RMS	5.92	5.98	7.77	6.05	6.35	6.33	31.4	31.7	32.0
AC RMS	5.90	5.97	7.56	1.96	2.56	2.23	20.4	21.14	20.97
Ripples	99.67%	99.99%	97.3%	32.4%	40.31%	35.23%	64.97%	66.7%	65.53%

Table 8 Experimental result comparison at 100% load

	Phase current (A)			DC link current (A)			Phase voltage (V)		
	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3
Peak to peak	52	72	74	10.1	8.4	14.1	50	50	50
Average	0.14	0.19	0.37	19.71	20.21	19.82	25.98	23.4	23.28
DC RMS	16.8	16.6	16.8	19.6	20.3	20.0	32.4	30.5	30.6
AC RMS	16.8	16.6	16.8	1.75	1.89	2.41	19.34	19.54	19.84
Ripples	100%	100%	100%	8.93%	9.31%	12.06%	59.69%	64.06%	64.84%

operating condition. DC RMS current is also minimal for the proposed method.

Mechanical speed waveforms are estimated from oscilloscope data and are graphically filtered, and results presented in Figs. 20 and 21 for normal speed curves and transient response, respectively. Figure 20a shows the speed ramp curve for 30% load condition and Fig. 20b for 100% load. The motor speed ramping starts after initial boundary condition checks and bootstrapping in controller, with initial push at $t = 0.6$ s.

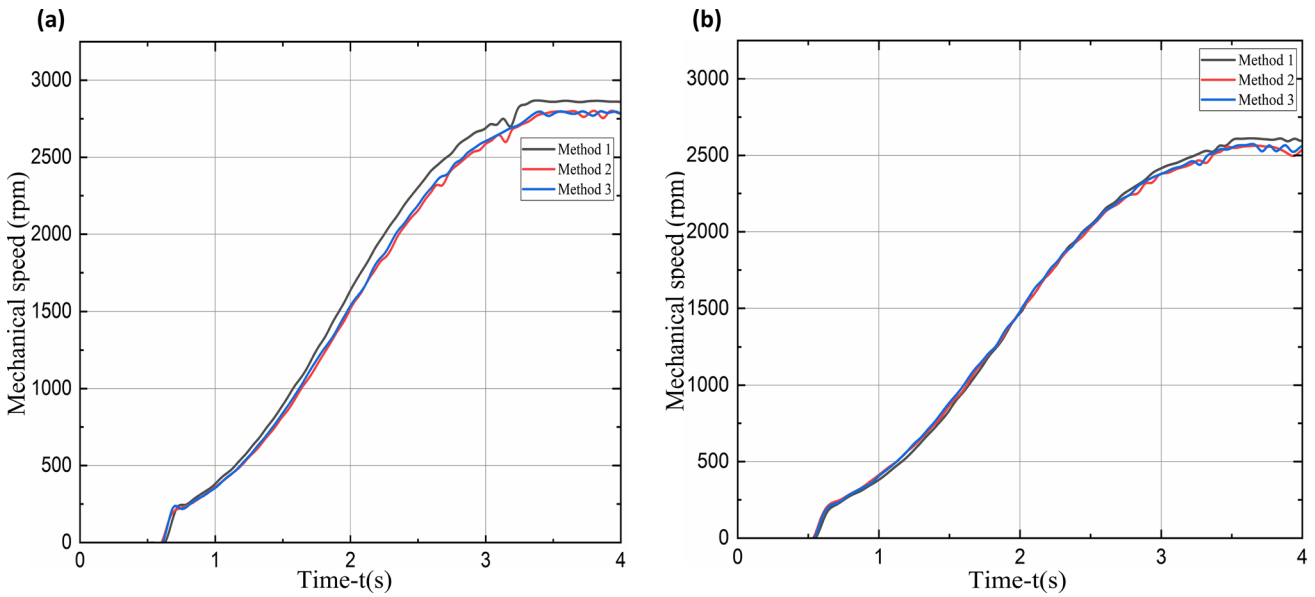


Fig. 20 Estimated mechanical speed for all three methods: **a** 30% loading; **b** 100% loading

It is observed that Method 1 obtains the steady state speed more accurately as compared to other two methods. The final speed difference between Method 1 and Method 2 at 30% load condition is observed as 60 RPM, and at full load it is observed as 50 RPM.

Transient response of the system is checked with all three methods by sudden full load application and estimated speed curves are presented in Fig. 21a and b. The response of method 1 is nearly same with steady speed decline after $t = 1.25$ s. Method 2 exhibits larger speed dip at $t = 1.2$ s. Method 1 exhibits fastest approach toward steady slope of speed decrease as shown in Fig. 21b.

Efficiency and torque per ampere are compared with Method 2 without any FTC taken as unit base for comparison and presented in Table 9. Method 1 exhibits highest efficiency with increment ranging from 0.8% to 2.2% and nominal torque/ampere increment varying from 2 to 9%. Figure 22a presents the radar diagram for combined comparison based on phase current, DC current ripples and memory execution as parameters. Proposed method has minimum area and represents optimum strategy to achieve best performance level. Figure 22b radar diagram for combined comparison is based on efficiency and torque per ampere. Proposed method 1 with maximum area again represents optimum strategy based on selected parameters.

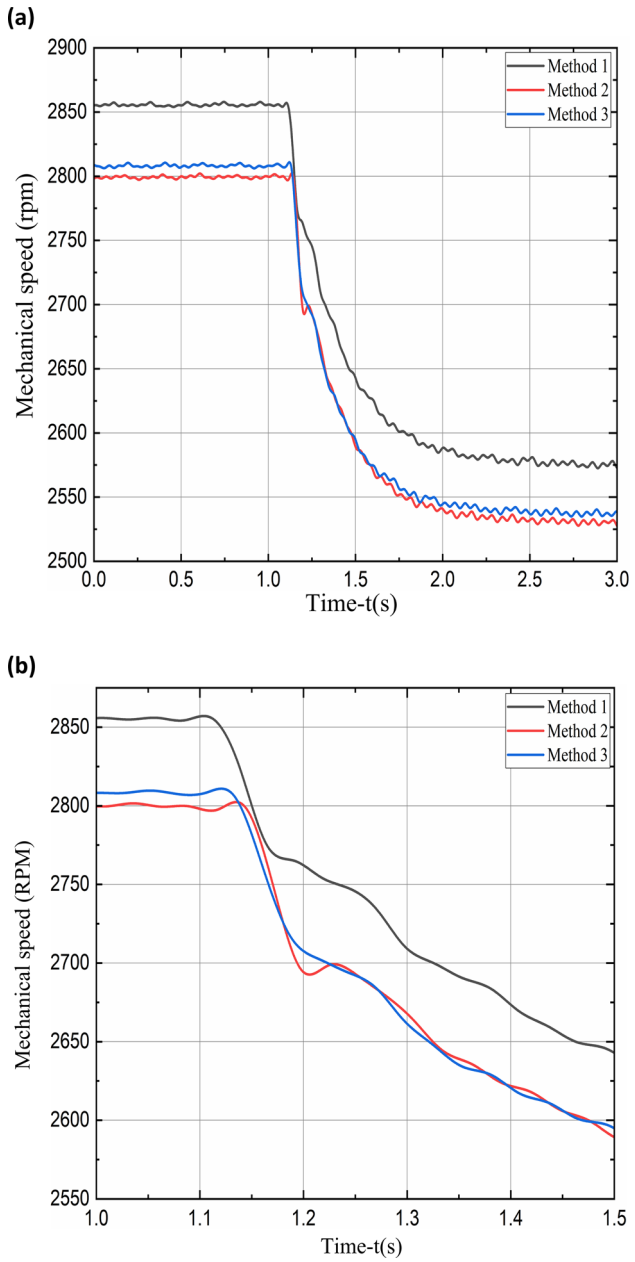


Fig. 21 Estimated mechanical speed curves for sudden load application: **a** full speed change; **b** transient response

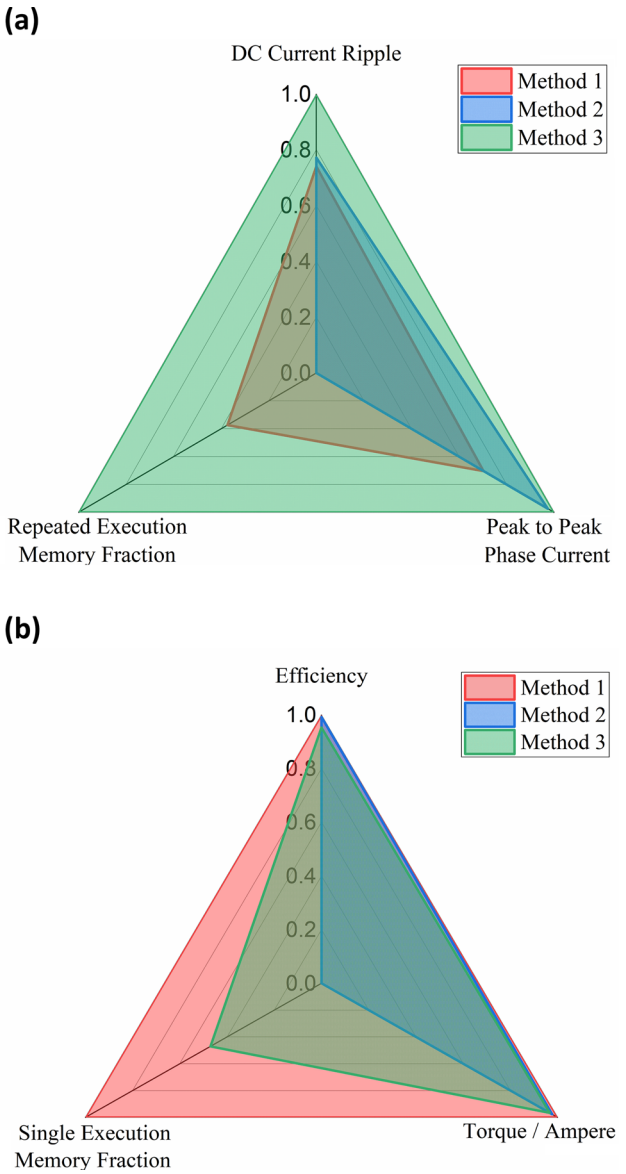


Fig. 22 Per-unit radar diagram for experimental result comparison at steady state: **a** minimum value parameters; **b** maximum value parameters

Table 9 Normalized efficiency and torque per ampere

	Efficiency scaling			Normalized torque/ampere		
Method	1	2	3	1	2	3
100% load	1.008	1	0.96	1.025	1	0.99
30% load	1.015	1	0.93	1.014	1	0.95
No load-50% PWM duty	1.022	1	0.91	1.096	1	0.94

7 Conclusions

This paper presented a scheme for controlling the effects of unbalanced hall sensor signals in sensored BLDC motor drives. Complete FTC scheme, integration in controller and comparison results with another method applicable in same type of faults are presented. The results of simulation studies and experimental tests obtained from the test bench of 48 V, 1 kW BLDC motor generally used in electric rickshaw applications, verified that the proposed scheme improves steady state and dynamic performance of the drive. Results show improvement in motor phase current, phase voltage and DC link current profiles with reduced peak to peak values. The proposed scheme has significantly less ripple content in stator phase and DC link currents at different loads. Additional memory loading is also less when compared with another LSM based scheme.

One important outcome of this FTC mechanism is reduced heating of controller and fewer battery discharge/charge cycles due to reduction in peak to peak and RMS values of phase currents and DC link current. This leads to better efficiency of the motor controller system and increases the possibility of battery life enhancement with better power utilization without any cost of additional hardware. The proposed scheme may be useful for automotive applications such as two wheelers and electric rickshaw which are cost sensitive, and other applications where controller circuit modification for hall sensor signal faults is not feasible.

Author contributions All authors contributed to the concept and manuscript preparation. The first draft of the manuscript was written by AG and all authors contributed to manuscript finalization. Research set up, data collection and analysis were performed by AG and MS. All authors read and approved the final manuscript.

Funding Not applicable.

Data availability Not applicable.

Declarations

Conflict of interest No competing interests. The authors have no competing interests as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Ethical approval Not applicable.

References

1. Mousmi A, Abbou A, El Houm Y (2020) Binary diagnosis of hall effect sensors in brushless DC motor drives. IEEE Trans Power Electron 35(4):3859–3868. https://doi.org/10.1109/TPEL.2019.2934794
2. Kolano K (2019) Improved sensor control method for BLDC motors. IEEE Access 7:186158–186166. https://doi.org/10.1109/ACCESS.2019.2960580
3. Alaeinovin P, Chiniforoosh S, Jatskevich J (2008) Evaluating misalignment of hall sensors in brushless DC motors. In: 2008 IEEE Canada electric power conference IEEE 1-6. https://doi.org/10.1109/EPC.2008.4763350
4. Park JS, Choi JH, Lee J (2016) Compensation method of position signal error with misaligned hall-effect sensors of BLDC motor. J Electr Eng Technol 11(4):889–897
5. Goswami A, Sreejeth M, Singh M (2022) Simulation and analysis of hall sensor misalignment in BLDC motor drive. In: 2022 IEEE Delhi section conference (DELCON), New Delhi, India. IEEE, 1–6. https://doi.org/10.1109/DELCON54057.2022.9753351
6. Beran P, Stahl-Offergeld M, Peters V, Krause D, Hohe H (2019) Impact of contact misalignment on magnetic cross sensitivity of integrated vertical hall sensors. IEEE Trans Magn 55(1):1–4
7. Aqil M, Hur J (2020) A direct redundancy approach to fault-tolerant control of BLDC motor with a damaged hall-effect sensor. IEEE Trans Power Electron 35(2):1732–1741. https://doi.org/10.1109/TPEL.2019.2917559
8. Zandi O, Poshtan J (2019) Fault diagnosis of brushless DC motors using built-in hall sensors. IEEE Sens J 19(18):8183–8190
9. Miguel-Espinar C, Heredero-Peris D, Igor-Gross G, Llonch-Masachs M, Montesinos-Miracle D (2022) Accurate angle representation from misplaced hall-effect switch sensors for low-cost electric vehicle applications. IEEE Trans Ind Appl 58(4):5227–5237. https://doi.org/10.1109/TIA.2022.3170920
10. Xi Z, Zhao Y, Lin H, Saleem R, Hassan E (2021) Real-time fault diagnosis and fault-tolerant control strategy for hall sensors in permanent magnet brushless DC motor drives. Electronics 10(11):1268. https://doi.org/10.3390/electronics10111268
11. Park DH, Nguyen AT, Lee DC, Lee HG (2017) Compensation of misalignment effect of hall sensors for BLDC motor drives. In: 2017 IEEE 3rd international future energy electronics conference and ECCE Asia (IFEEC 2017-ECCE Asia), Kaohsiung, Taiwan. IEEE, 1659–1664
12. Ebadpour M, Amiri N, Jatskevich J (2021) Fast fault-tolerant control for improved dynamic performance of hall-sensor-controlled brushless DC motor drives. IEEE Trans Power Electron 36(12):14051–14061. https://doi.org/10.1109/TPEL.2021.3084921
13. Aqil M, Hur J (2021) Multiple sensor fault detection algorithm for fault tolerant control of BLDC motor. Electronics 10(9):1038
14. Scelba G, De Donato G, Pulvirenti M, Giulii Capponi F, Scarcella G (2016) Hall-effect sensor fault detection, identification, and compensation in brushless DC drives. IEEE Trans Ind Appl 52(2):1542–1554. https://doi.org/10.1109/TIA.2015.2506139
15. Park JS, Lee KD (2017) Online advanced angle adjustment method for sinusoidal BLDC motors with misaligned hall sensors. IEEE Trans Power Electron 32(11):8247–8253
16. Microelectronics ST (2019) 1 kW traction controller reference design for 3-phase BLDC motors for light electric vehicles. DB4061, STDES-EVT001V1, Rev. 1. https://www.st.com/resource/en/data_brief/stdesevt001v1.pdf. Accessed Feb 2020
17. Microelectronics ST (2019) How to use the 1 kW traction controller reference design for 3-phase BLDC motors for light electric vehicles. UM2647, Rev. 1, User Manual. https://www.st.com/resource/en/user_manual/um2647-how-to-use-the-1-kw-traction-controller-reference-design-for-3phase-bldc-motors-for-light-electric-vehicles-stmicroelectronics.pdf. Accessed 20 Novemb 2021
18. Kolano K, Moradewicz AJ, Drzymała B, Gęca J (2022) Influence of the placement accuracy of the brushless DC motor hall sensor on inverter transistor losses. Energies 15(5):1822

19. Krishnan R (2012) Permanent-magnet synchronous and brushless DC motor drives. In: Prentice-Hall U (ed) Electric motor drives: modeling, analysis, and control, Indian. Saddle River, New Jersey, pp 513–620
20. Nama T, Gogoi A K, Tripathy P (2017) Application of a smart hall effect sensor system for 3-phase BLDC drives. In: 2017 IEEE international symposium on robotics and intelligent sensors (IRIS), Ottawa, ON, Canada. IEEE, 208–212. https://doi.org/10.1109/IRIS.2017.8250123
21. Goswami A, Sreejeth M, Upadhyay R, Indu S (2021) Brushless DC motor based e-rickshaw controller design. In: 2021 international conference on recent trends on electronics, information, communication & technology (RTEICT), Bangalore, India. IEEE, 196–200 https://doi.org/10.1109/RTEICT52294.2021.9573560
22. Designs TI (2015) Automotive 1-kW 48-V BLDC Motor drive reference design. Texas instruments, Inc. https://www.ti.com/lit/pdf/tiduay9. Accessed 26 Jan 2022

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

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

Journal of Electronic Materials
https://doi.org/10.1007/s11664-024-10948-7

ORIGINAL RESEARCH ARTICLE

Investigation of a Gate Stack Gate-All-Around Junctionless Nanowire Field-Effect Transistor for Oxygen Gas Sensing

Rishu Chaujar[1] · **Mekonnen Getnet Yirak**[1,2]

Received: 23 September 2023 / Accepted: 16 January 2024
© The Minerals, Metals & Materials Society 2024

Abstract

The design and analysis of a gate stack gate-all-around junctionless nanowire field-effect transistor (GAA-JL-NWFET) with a catalytic metal as gate contact for oxygen gas sensing is presented in this study. An n-channel GAA-JL-NWFET design using gold (Au) as the gate metal electrode is employed for oxygen sensing by utilizing appropriate work function values, which interact with oxygen gas and change the device's electrical properties. This work focuses on changes in temperature (300–500 K) and Au metal gate work function (5.05–5.20 eV) to investigate the presence of oxygen molecules and their impact on the GAA-JL-NWFET gas sensor performance. Changes in the surface potential, threshold voltage, hole concentration, electron concentration, subthreshold voltage, electric field, and drain current using the ATLAS TCAD simulator are investigated for the adsorption of gas molecules to determine the electrical characteristics of the proposed device. Changes in threshold voltage (V_{th}), switching ratio, and subthreshold current sensitivity ($S_{I_{OFF}}$) can be considered as sensitivity parameters for sensing oxygen gas molecules. The results reveal that as the Au work function shifts at the gate by 100 mV, the sensitivity ($S_{I_{OFF}}$) enhancement using gate stack GAA-JL-NWFET-based oxygen gas sensors compared to GAA-MOSFET and conventional MOSFET are 15.13% and 88.31%, respectively. Based on our simulation results, the proposed device offers excellent sensitivity, low power consumption, and a fast response time, making it an appropriate candidate for oxygen gas sensing, including environmental monitoring, medical diagnosis, and industrial safety.

Keywords Junctionless · oxygen gas sensor · gate-all-around · gate stack · nanowire field-effect transistor

Introduction

For the last two decades, field-effect transistors (FETs) have been widely used to develop logic gates and other integrated circuits.[1] However, designing and developing low-dimensional nano-electronic devices for next-generation sensing primitives requires semiconducting materials that can be scaled down to the atomic scale. Recently, ultralow-power-consuming devices that meet nanoscale manufacturing, high sensitivity, and high package density requirements and are also technologically feasible are being investigated by the rapidly developing very-large-scale integration (VLSI) industry.[2] These sensor devices have garnered considerable attention in the scientific community due to their ability to accurately detect a wide variety of gas molecules at ambient temperature.[3] Metal–oxide–semiconductor field-effect transistor (MOSFET)-based gas sensors meet the growing need for gas detection in automotive, safety, and medical industries and in environmental monitoring due to their low power, low cost, high sensitivity, small size, and complementary metal–oxide–semiconductor (CMOS) compatibility with microelectronics fabrication processes.[4–6] A gas sensor based on a MOSFET was recently developed to detect gas species by measuring the induced change in work function at the film's surface.[3] It is important to note that the change in the work function of the metal (gate) caused by the adsorption of gas molecules is the fundamental principle behind FET-based gas sensors.[7] For instance, dipole formation causes a change in the work function of the metal electrode.

In addition, the threshold voltage and subthreshold current of the FET also change significantly due to a change

✉ Rishu Chaujar
chaujar.rishu@dtu.ac.in

Mekonnen Getnet Yirak
mekonnengetnet01@gmail.com

[1] Department of Applied Physics, Delhi Technological University, New Delhi, Delhi, India

[2] Physics Department, Debre Tabor University, Debre Tabor, Ethiopia

Published online: 17 February 2024

in the metal's work function. The change in the metal work function also has an effect on the FET's threshold voltage, which shifts significantly. Several gas sensor models have been reported in the literature in recent years, emphasizing various FET architectures.[8,9] As a gas sensor, a gate stack gate-all-around junctionless nanowire field-effect transistor (GAA-JL-NWFET) was proposed based on a catalytic metal gate. These sensors promise to be some of the simplest, least expensive, and most effective technologies for real-time measurement and gas leakage detection.[10] While planar MOSFET [11] and tunnel field-effect transistor (TFET) gas sensors[12] have shown great promise, they face scaling challenges at extremely small sizes. A high surface-to-volume ratio is required to enhance sensitivity via enhanced surface reactions.[13] Because of its high surface-to-volume ratio, practically optimal subthreshold properties, superior gate control, and low leakage current,[5,13] the GAA nanowire MOSFET is a promising device for the development of a low-power, sensitive, and nanoscale CMOS-compatible gas sensor. Previously reported work[4–6,14] details the development of FET-based gas sensors with a catalytic metal gate. Recently, a silver catalytic metal gate FET-based oxygen gas sensor was introduced.[15] However, no studies have been conducted on cylindrical gate stack GAA-JL-NWFETs with catalytic metal (Au) gate electrodes for oxygen gas detection applications.

This work proposes a gate stack GAA-JL-NWFET-based oxygen gas sensor with a catalytic metal (Au) gate for the first time to enhance device sensitivity. The gate stack GAA device structure has various advantages over conventional FETs, such as better gate control and reduced short-channel effects. Gas molecules react at the catalytic metal (Au) gate surface, changing its work function and the MOSFET device's flat-band voltage, threshold voltage, and drain current.[13,16] This reaction is crucial because shifting the gate metal's work function has an essential effect on modulating the device's electrical properties.[8] This shift in conductivity is proportional to the amount of a definite gas that exists in the environment, which enables a quantitative assessment of the presence of oxygen gas and its concentration. When the metal (Au) gate is exposed to the O_2 environment, the oxygen gas molecules break apart at the metal surface and diffuse into the metal gate[16] due to the absorption of oxygen gas by a catalytic metal gate (Au), leading to the formation of gold trioxide.[16]

Furthermore, due to the adsorption of the gas molecules into the gate metal, an electrical dipole layer is formed at the Au/HfO_2 interface, and the metal work function changes appropriately. In this scenario, we assessed the proposed device's surface potential distribution and electric field in the channel, threshold voltage, subthreshold voltage, switching ratio, drain current, and device sensitivity with respect to shifts in threshold voltage and subthreshold current. In order to develop innovative and high-performance electrical devices at the nanoscale, researchers have opted for high-κ gate oxide, such as hafnium oxide (HfO_2) and interface oxide (SiO_2).[17–19] When compared to other higher-dielectric materials,[1,17] hafnium oxide (HfO_2) stands out as the best and most potent option for enhancing sensing performance in silicon nanowire channels.[1,17,20] Inter-diffusion between the channel and the gate electrode is prevented by incorporating the interface layer.[20] In this work, we developed an n-substrate GAA-JL-NWFET-based gas sensor to investigate its performance for oxygen gas sensing using the ATLAS 3D TCAD (technology computer-aided design) device simulator.

Device Physical Structure

Figure 1 shows a 3D description and 2D cross-section of the cylindrical GAA-JL-NWFET-based oxygen gas sensor. Here, L (20 nm) is the length of the dielectric material (HfO_2); (SiO_2) is an interface oxide layer; T_3, T_2, and T_1 are the thicknesses of interface oxide, hafnium oxide, and metal gate of the proposed device, respectively; and

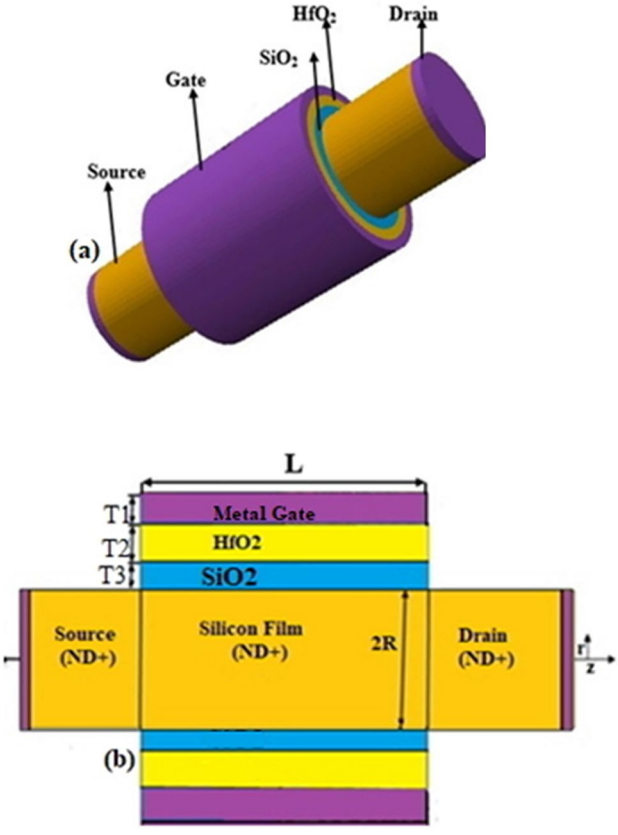


Fig. 1 (a) 3D description and (b) 2D cross-sectional interpretation of cylindrical GAA-JL-NWFET-based oxygen gas sensor.

2R is the diameter of the channel (silicon film thickness). The device gate length is 20 nm for all simulations, and the source and drain lengths are 10 nm each, as indicated in Table I. To produce a sustainable threshold voltage in the n-substrate channel devices that we designed, we applied a high doping concentration ($N_{D+} = 1 \times 10^{19}\text{cm}^{-3}$) all the way across the channel, from the source to the drain. The supply gate–source voltage (0.6 V) with a constant drain–source voltage (0.05 V) was used for all simulations.

Simulation Models

All simulations in this work are performed using the Silvaco ATLAS TCAD tool. To incorporate electron mobility models, drift-diffusion, concentration-dependent, and field-dependent mobility models are activated.[26] The driving current due to the charge carrier movement is also taken into consideration by the drift-diffusion model. The Shockley–Read–Hall (SRH) model accounts for majority–minority charge carrier recombination.[26] Fermi–Dirac statistics are introduced due to heavy source and drain doping. Both parallel and perpendicular field mobility are accounted for using the Boltzmann transport statistics and the Lombardi mobility model.[27] A gold (Au) gate metal with a work function (Φ_m) of 5.00 − 5.20 eV is employed on top of a high-κ oxide stack to detect the oxygen gas molecules. Furthermore, 5-nm-thick Si NW film has been chosen as the channel material, with a channel length of 20 nm. The effect of temperature on the sensitivity of oxygen gas sensors is examined. The sensitivity of the gate stack GAA-JL-NWFET is compared to that of the bulk MOSFET and GAA MOSFET, which are used as FET-based gas sensors.

Figure 1 depicts the three-dimensional construction and cross-sectional view of the n-substrate metal (Au) gate GAA-JL-NWFET-based oxygen gas sensor. In this work, the catalytic metal gate approach was considered to describe the operation of a GAA-JL-NWFET-based oxygen gas sensor. The experimental data are used to calibrate the simulation's model parameters.[28] Quantum effects cannot be considered because the radius of a silicon pillar is not less than 5 nm.[16]

Validating simulation results against experimental data is necessary to ensure that the simulation model accurately replicates the physical behavior of the gate stack GAA-JL-NWFET.[2,29,30] By validating the model, one can ensure that the simulation parameters, approximations, and assumptions are reliable and accurate. Comparison enables scientists and engineers to fully understand the real-world performance of the gate stack GAA-JL-NWFET. Experimental results give information about the device's real behavior by considering variables that the simulation model might not have fully included.

The gold work function must play an essential role in changing the electrical field characteristics of the device.[31] As gas molecules react with the metal surface, the work function of the catalytic metal gate changes. Also, the impact of temperature variations (300–500 k) on device performance has been examined. A shift in the device's threshold voltage and drain current indicates its sensitivity to gas. As a result, we have examined improved device performance in electrical characteristics such as switching ($I_{\text{ON}}/I_{\text{OFF}}$) ratio and subthreshold current sensitivity ($S_{I_{\text{OFF}}}$). The choice of parameters used in modeling and simulation is validated by the proximity of the simulated results to the experimental results,[25] as illustrated in Fig. 2.

Results and Discussion

The field-effect transistor (FET) is a transducer that converts the shift in work function at the surface of a sensitive film or catalytic metal into the appropriate electrical signal.[5,32] A shift in flat-band voltage causes a change in the surface potential, subthreshold current, and threshold voltage (V_{th}).

Table I Technology parameters for device design

Device parameters	Gate stack GAA-JL-NWFET
Size of the channel (nm)	20.0
Thickness of SiO_2& HfO_2 , respectively (nm)	0.3 and 1.5
Thickness of the interface oxide (SiO_2) (nm)	1.0
Length of silicon dioxide (SiO_2)(nm)	20.0
Length/thickness of drain/source (nm)	10.0
Hafnium oxide length (nm)	20.0
Silicon film radius (nm)	10.0
Channel, source, and drain doping (N_D+)	10^{19}cm^{-3}[21]
Oxide dielectric constant,HfO_2 and SiO_2	25.0 and 3.90[22–24]
Reference gate work function, gold (Au), (eV)	5.0[25]

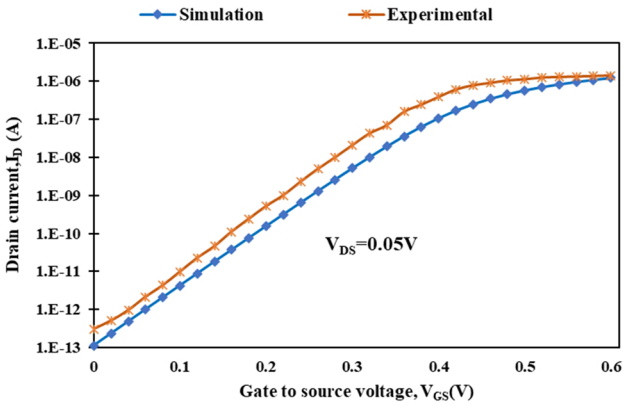


Fig. 2 Calibration of simulation results at $V_{DS} = 0.05$ V with experimental results[28] for the gate stack GAA-JL-NWFET oxygen sensor.

Drain Current

The drain current of an oxygen gas sensor can be altered by a variety of factors, including a gold electrode work function and temperature variations, as illustrated in Fig. 3. The reactivity of oxygen gas molecules at the catalytic metal gate surface induces a change in the work function of the gate metal, which causes a shift in flat-band voltage due to additional band bending. The oxygen molecules may donate or accept electrons from the gold surface during their interaction. This charge transfer alters the device's conductivity by changing the potential barrier at the metal–semiconductor interface. The change in the work function of the gold electrode caused by oxygen adsorption can alter the barrier height and affect the current flow in the sensor, as shown in Fig. 3a. As the work function of the gold electrode increases due to oxygen adsorption, the drain current decreases, and as the work function decreases, the drain current increases. This is due to higher work functions increasing the barrier height, which reduces the flow of electrons between the electrode and gas, decreasing the drain current.

In contrast, a reduced work function lowers the barrier height, increasing electron flow and drain current. In summary, differences in the gold work function change the barrier height between the gold electrode and the surrounding gas, affecting the oxygen gas sensor drain current. Increases in work function reduce drain current, while decreases increase it. In addition to the catalytic metal gate work function, the electrical conductivity of the detecting material is affected by temperature, which influences the sensor's sensitivity to oxygen gas, as shown in Fig. 3b. The chemical reaction rate between oxygen molecules and the sensing material typically accelerates as the temperature rises. This can increase the drain current. This behavior is explained by the fact that higher temperatures supply more thermal energy

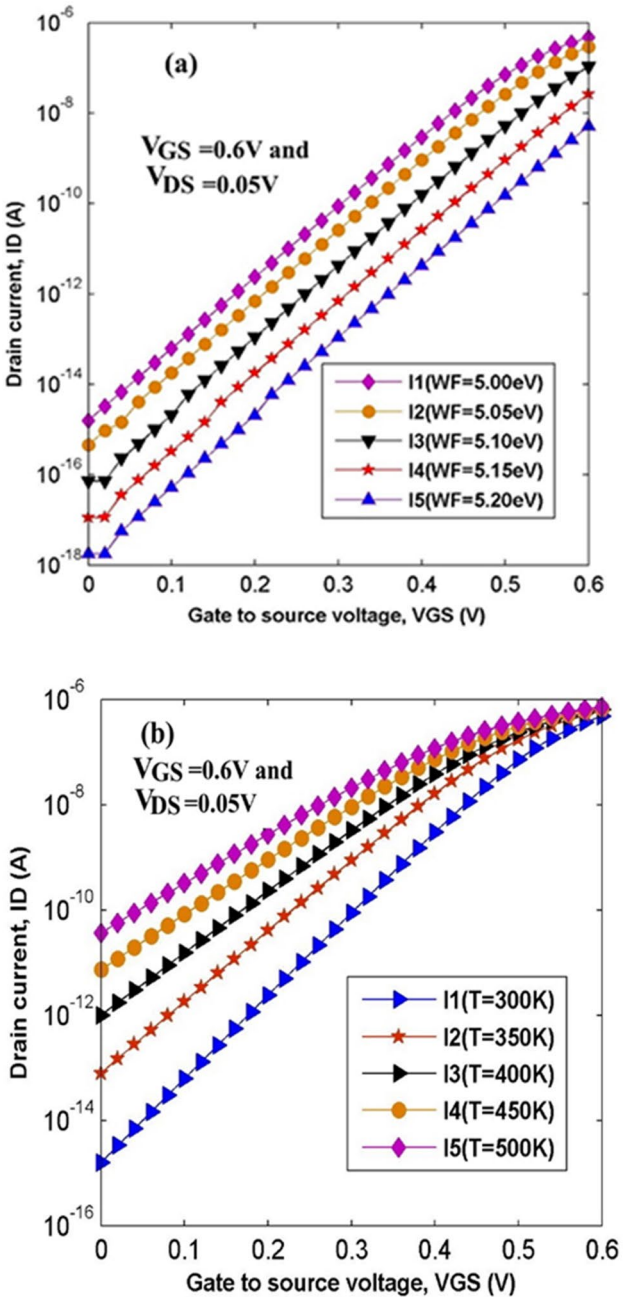


Fig. 3 Effect of (a) gold (Au) work function and (b) temperature on I_D–V_G for a cylindrical gate stack GAA-JL-NWFET.

to the reacting species, allowing them to overcome energy barriers and permit faster reactions.

Surface Potential

Variations in gold work function and temperature can affect the surface potential of a gate stack GAA-JL-NWFET oxygen gas sensor, as shown in Fig. 4. The potential distribution along the channel for various (Φ_m)

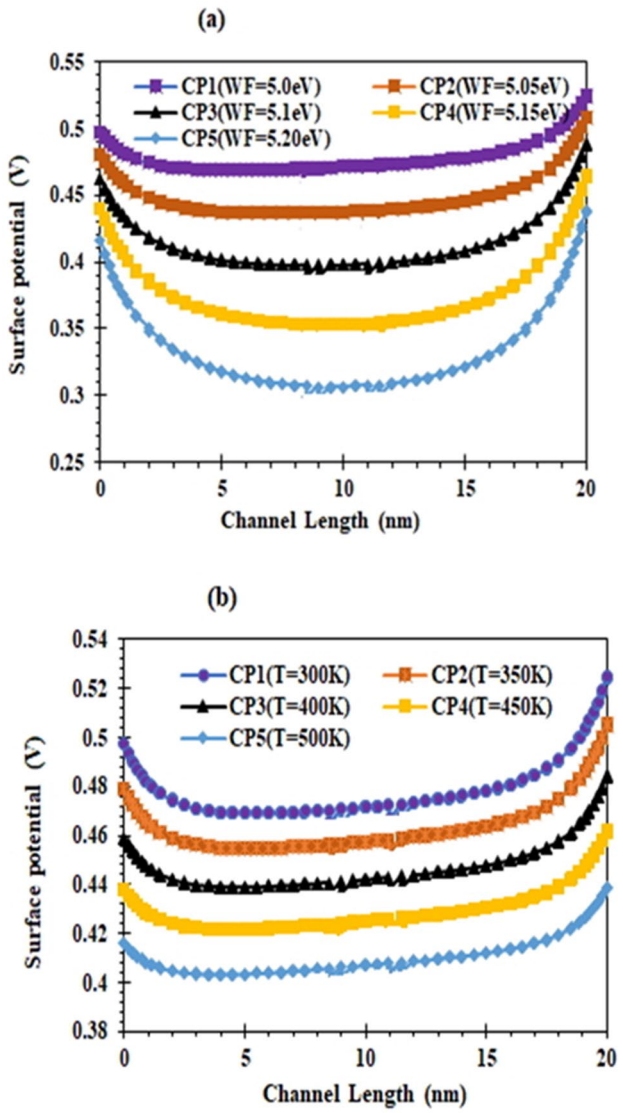


Fig. 4 Effect of (a) gold (Au) work function and (b) temperature on surface potential (V) for cylindrical gate stack GAA-JL-NWFET (V_{GS} = 0.6 V and V_{DS} = 0.05 V).

values are shown in Fig. 4a. Changes in the gold work function can change the potential barrier at the gold/semiconductor interface by altering the contact potential difference. Temperature changes influence the materials' conductivity and the charge carriers' energy distribution, resulting in surface potential change, as shown in Fig. 4b.

Increased metal gate electrode work function raises the metal/semiconductor interface potential barrier for an *n*-type MOSFET, as shown in Fig. 4a. This increased potential barrier slows electron transport from the source to the drain, leading to decreased channel potential.

Electric Field

Variations in the gold work function influence the contact potential difference between the gold gate electrode and the semiconductor channel material. This variation in the contact potential difference modifies the electrostatic potential distribution along the channel and influences the electric field, as illustrated in Fig. 5a. As clearly stated, a higher gold electrode work function variation in *n*-type MOSFETs leads to an increased electric field along the channel due to a raised potential barrier at the gold/semiconductor interface, as higher barrier potential attracts and confines electrons to the channel surface. On the other hand, a weaker electric field in the channel region results from a reduced gold work function, as the potential barrier at the interface is lowered. In addition, the conductivity of materials, including the semiconductor channel and the gold gate electrode, increases with increasing temperature, as illustrated in Fig. 5b. Because temperature affects the electrical conductivity of materials, the energy distribution of charge carriers leads to reduced electric fields at higher temperatures due to improved conductivity, reducing potential barriers within the device.

Device Sensitivity Parameters (Threshold Voltage and Leakage Current)

A larger gold work function increases the potential barrier at the metal/semiconductor interface, which raises the threshold voltage of an *n*-type device to induce current flow, as depicted in Fig. 6a. In addition to a metal electrode work function, temperature variations influence the concentration and mobility of carriers in the semiconductor channel, which influences the threshold voltage, as depicted in Fig. 6b. As the temperature rises, the carrier concentration also increases, resulting in lower threshold voltage. This is because more carriers are available to conduct current, requiring less voltage to reach the desired level of current flow.

The sensitivity of the *n*-type GAA-JL-NWFET oxygen gas sensor can be determined by analyzing the shift in threshold voltage caused by the gold work function and temperature variations, as shown in Fig. 7. The potential barrier at the metal/semiconductor interface is increased when the work function of gold is increased, leading to a higher threshold voltage shift. Higher sensitivity will be indicated by a higher threshold voltage shift, as shown in Fig. 7a. This clearly indicates that the difference in work function between the gate electrode and the semiconductor channel is proportional to the change in threshold voltage due to this variation. In addition, the charge carrier concentration and mobility within the semiconductor channel are both impacted by temperature, affecting the shift in

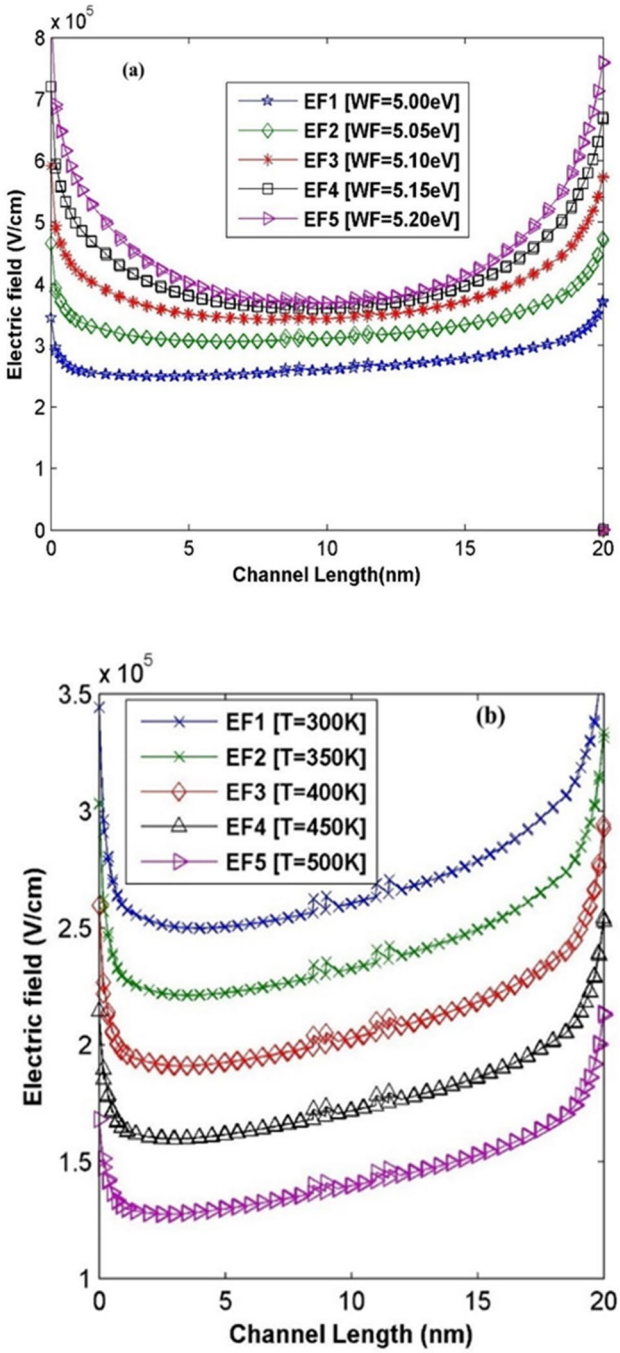


Fig. 5 Impact of (a) gold (Au) work function and (b) temperature on the electric field (V/cm) for a cylindrical gate stack GAA-JL-NWFET, (V_{GS} = 0.6 V and V_{DS} = 0.05 V).

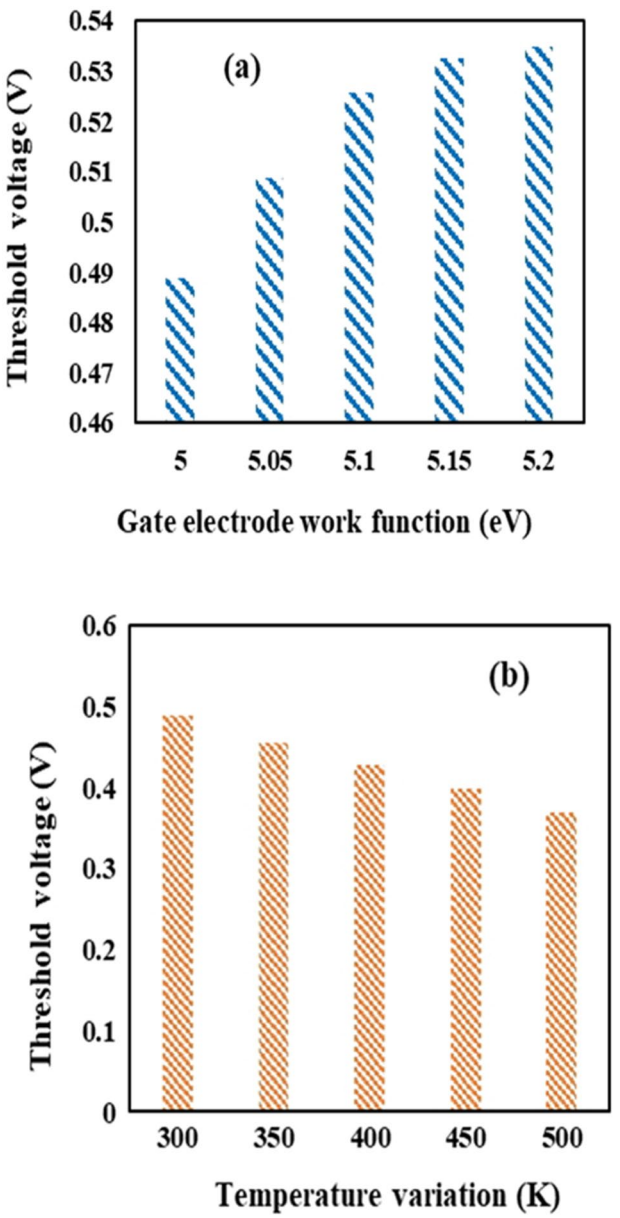


Fig. 6 Effect of (a) gold (Au) work function and (b) temperature on threshold voltage (V_{th}) for the cylindrical gate stack GAA-JL-NWFET (V_{GS} = 0.6 V and V_{DS} = 0.05 V).

threshold voltage, as depicted in Fig. 7b. Thus, the presence of oxygen gas molecules can be detected using a gold catalytic metal gate electrode by observing the shift in V_{th}. Higher threshold voltage shifts imply higher sensitivity to changes generated by the target gas (oxygen), whereas more minor shifts indicate lower sensitivity.

As the temperature and work function of the palladium metal increase, a change in the threshold voltage (ΔV_{th}) arises due to the increase in the concentration of hydrogen gas molecules, as shown in Fig. 7, which can be calculated using Eq. 1.[6]

$$\Delta V_{th} = \left| V_{th(\text{after gas reaction})} - V_{th(\text{before gas reaction})} \right| \quad (1)$$

Another measure of device sensitivity is the switching ratio (I_{ON}/I_{OFF}) of an *n*-type GAA-JL-NWFET oxygen gas sensor, which can change with variations in the gold work

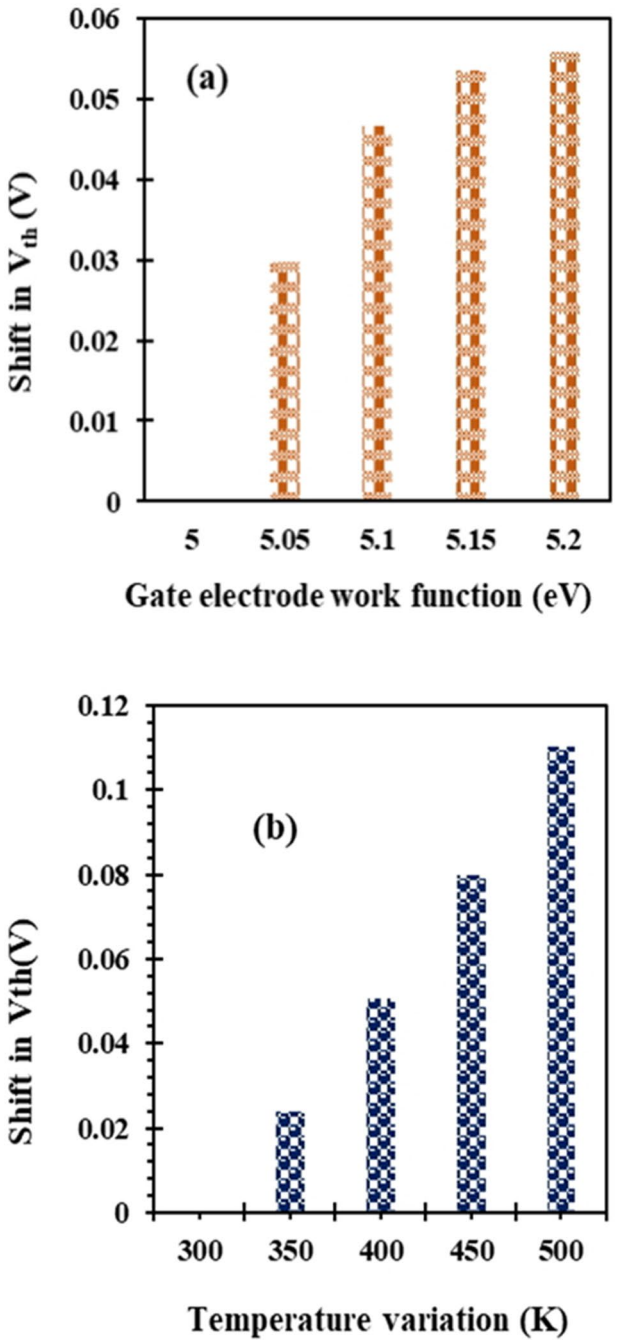


Fig. 7 Variations in threshold voltage (ΔV_{th}) (a) gold (Au) metal gate work function and (b) temperature for a cylindrical gate stack GAA-JL-NWFET (V_{GS} = 0.6 V and V_{DS} = 0.05 V).

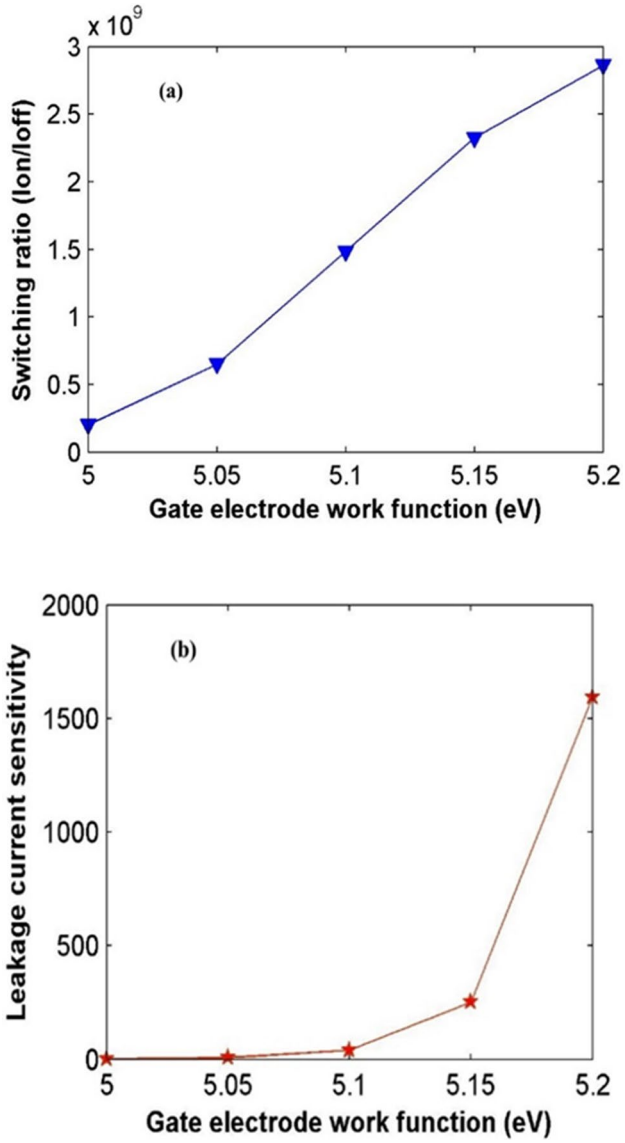


Fig. 8 Impact of the gold gate electrode on (a) switching ratio (I_{ON} /I_{OFF}) and (b) leakage current for a cylindrical gate stack GAA-JL-NWFET, (V_{GS} = 0.6 V and V_{DS} = 0.05 V).

function. The switching ratio is a crucial metric for measuring the sensitivity of the device.[33] The effect of variation in the gold electrode work function on the switching ratio of an n-channel gate stack GAA-JL-NWFET-based oxygen gas sensor is shown in Fig. 8a, and it has the potential to affect the switching ratio of the device. When the work function changes in millivolts, the switching and subthreshold current ratios shift exponentially.

Another metric measure of device sensitivity is the subthreshold current ratio ($S_{I_{OFF}}$), which can be impacted by variations in the gold work function and temperature.[5] The influence of changes in the work function of a catalytic metal (gold) electrode on the proposed device's subthreshold current sensitivity is studied to evaluate the device's overall performance and stability, as illustrated in Fig. 8b. It clearly describes ($S_{I_{OFF}}$) changes exponentially when the work function is changed in millivolts, indicating that the effect of gas molecules on the OFF-current is much greater than on the ON-current; thus, the subthreshold region offers much

higher sensitivity along with low-power operation, resulting in a low-cost gas sensing device.[5,13] This increased sensitivity in the subthreshold region can be traced to the additional band bending that takes place in the absence of Fermi level trapping due to the shift in the work function of the metal–semiconductor caused by gas molecule surface reactivity. It is essential to keep in mind that temperature can also affect the leakage current, which in turn has the potential to influence the OFF-state current ($S_{I_{OFF}}$) sensitivity as illustrated in Fig. 8b, which can be obtained using Eq. 2.[5,13,20]

$$S_{I_{OFF}} = \frac{I_{OFF(\text{after gas reaction})}}{I_{OFF(\textit{before}\ \text{gas reaction})}} \tag{2}$$

For instance, a change of 100 meV in the work function of a gold metal gate electrode for bulk MOSFET and gate stack GAA-JL-NWFET causes an 88.31% improvement in $S_{I_{OFF}}$; thus, the gate stack GAA-JL-NWFET has significantly higher sensitivity for oxygen gas detection than the bulk MOSFET.

To assess the overall performance and stability of the proposed device, the impact of temperature variations on the subthreshold current sensitivity is examined, as shown in Fig. 9. In this case, the leakage current sensitivity is affected by changes in the concentration of oxygen gas. Stated differently, it represents the transistor's ability to sense variations in the oxygen concentration of the surrounding air. In summary, there are several factors to consider when determining how temperature affects a gate stack GAA-JL-NWFET's leakage current sensitivity for oxygen gas detection. For the sensor to function reliably under various operating situations, its temperature-dependent behavior must be thoroughly examined and characterized. Creating a sensor that performs reliably and accurately in various temperature conditions may entail simulation research, experimental testing, and material and design parameter optimization.

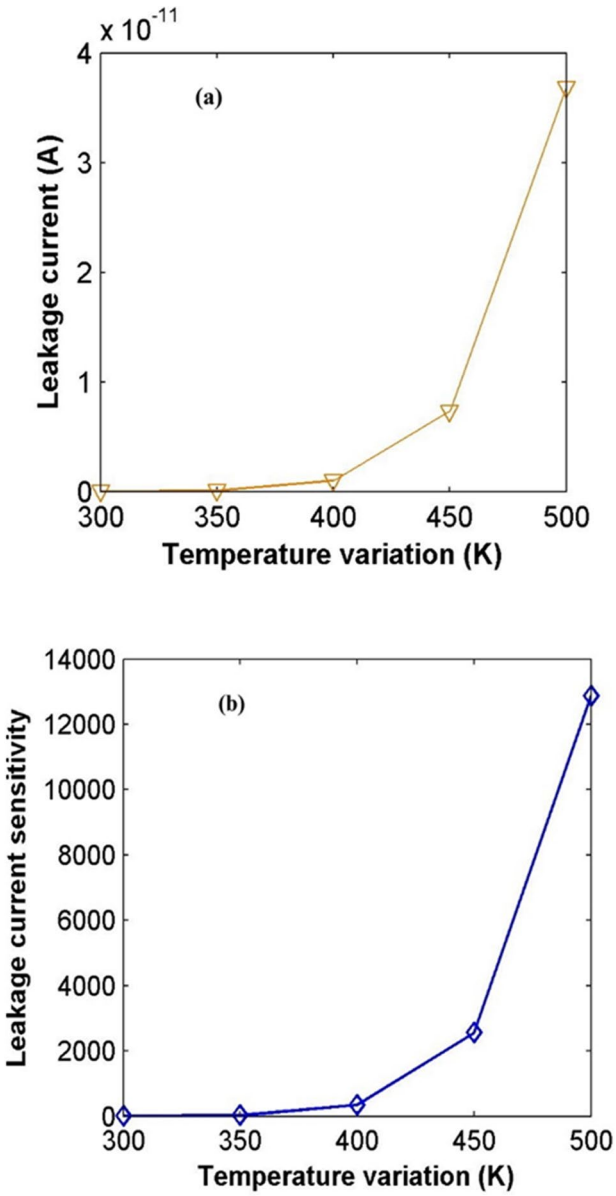


Fig. 9 (a) Leakage current and (b) leakage current sensitivity for a cylindrical gate stack GAA-JL-NWFET (V_{GS} = 0.6 V and V_{DS} = 0.05 V).

Stability and Reproducibility Evaluations

In the context of a gate stack GAA-JL-NWFET oxygen gas sensor, a transient solution typically refers to the system's reaction to a change or disruption in the oxygen concentration being monitored. The sensor's capacity to quickly and precisely identify changes in oxygen levels and to respond is frequently tied to transient solutions. These may include signal processing, data analysis, and calibration methods to ensure that the sensor responds to changes in oxygen concentration appropriately and quickly. Transient analyses with different gate electrodes were performed to evaluate the stability of the proposed sensor, as shown in Fig. 10a. The transient response was simulated by applying (V_{GS} = 1.2 V) a step time of 5×10^{-9} s, a ramp time of 1×10^{-9} s, and a stop time of 1×10^{-9} s. Figure 10a depicts the transient response of the drain current at different metal gate electrode work functions of the gate stack GAA-JL-NWFET oxygen gas sensor. The drain current for the GAA-JL-NWFET oxygen gas sensor was found to be greater at a lower electrode work function.

Also, the temperature has various effects on the sensor system, and understanding these effects is critical for maintaining dependable and accurate sensor performance. Temperature changes can cause baseline drift in sensor output. Baseline drift occurs when the sensor's output signal varies

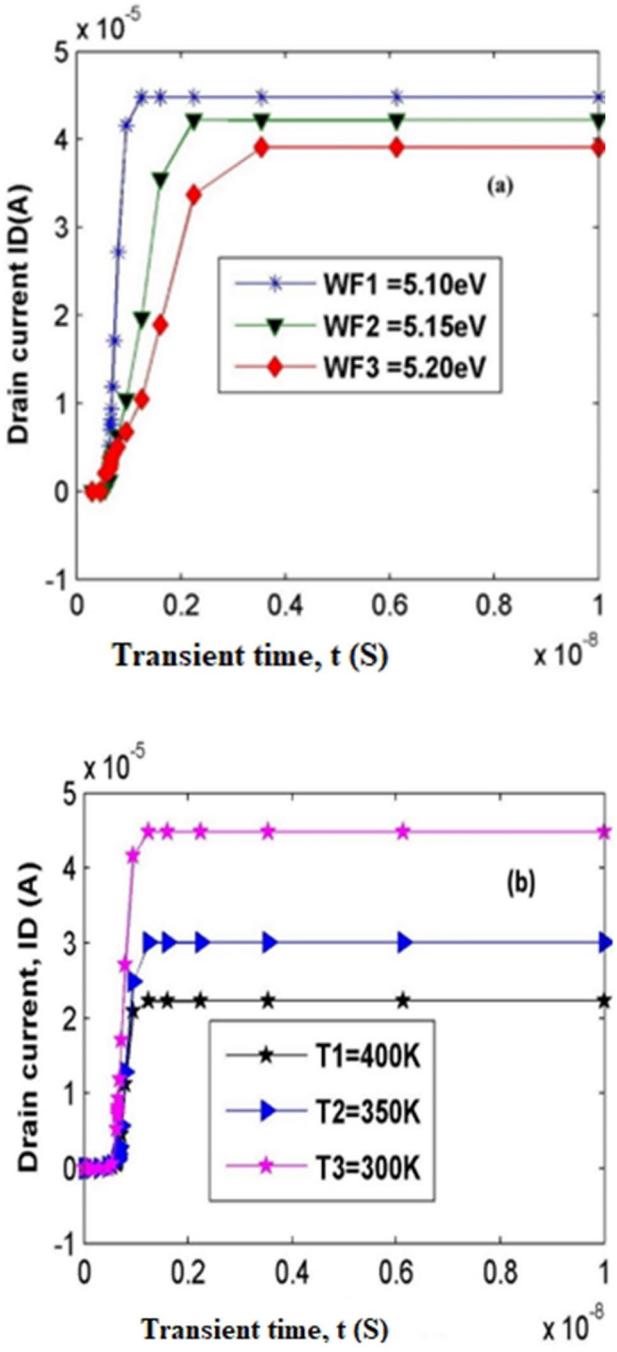


Fig. 10 Transient solution of the drain current for the (a) gold (Au) metal gate work function and (b) temperature for the cylindrical gate stack GAA-JL-NWFET (V_{GS} = 0.6 V and V_{DS} = 0.05 V).

even when no changes in the target gas (oxygen) occur. It is critical to keep baseline drift to a minimum in order to keep the sensor system accurate. The effect of temperature on the stability of transfer characteristics was investigated. Figure 10b shows the temperature dependence of the I_D–V_{GS} characteristics in the gate stack GAA-JL-NWFET oxygen gas sensor; the I_D–V_{GS} curves do not differ significantly between 300 K and 400 K. The drain current is measured for five different temperatures, and the results show that it is consistent, with negligible variance.

Long-term performance, consistency across sensors and batches, calibration reliability, and reaction to possible interferents are all evaluated as part of the stability and reproducibility assessments for the gate stack GAA-JL-NWFET oxygen gas sensors. These evaluations are necessary to guarantee the sensor's efficacy and dependability under various operating circumstances.

Table II presents a comparison of the sensitivity in terms of the order of shift in $S_{I_{OFF}}$, which is defined as the ratio of $S_{I_{OFF}}$ after and before the gas reaction at the gate metal. A comparison of the effects of work function and temperature change generated by gas molecules on $S_{I_{OFF}}$ for three different MOSFET architectures, namely bulk MOSFET, GAA MOSFET, and gate stack GAA-JL-NWFET, reveals some interesting results. When the work function of the gate metal is changed as a result of the reaction of gas molecules with a catalytic metal gate, the change in subthreshold current increases in the case of a gate stack cylindrical GAA-JL-NWFET because the channel is placed under the more effective control of the gate. Thinner silicon bodies are warped by interface and higher-dielectric oxides in addition to the catalatic metal gate, thus making it more sensitive because of higher surface-to-volume ratios, improved gate control, and lower subthreshold leakage current. GAA-JL-NWFET devices can be affected in multiple ways by using high-κ dielectric oxide HfO_2 (hafnium oxide) and interface oxide SiO_2 (silicon dioxide). HfO_2 has a higher dielectric constant (κ) than SiO_2, which increases the capacitance and channel gate control.[6,34,35] Because of its greater bandgap energy than SiO_2, this can improve the device's threshold voltage control, performance, reliability, and gate leakage current.[36,37] This enables the fabrication of smaller transistors, which enhances device density and performance while reducing power usage and increasing efficiency. Between the silicon substrate and the high-κ (HfO_2) gate dielectric, SiO_2 is often used as an interface oxide layer and used as a buffer layer, preventing direct contact between the high-κ material and the silicon substrate, reducing interface imperfections, and improving overall device performance. Since SiO_2 is an interface oxide, it can improve the dielectric-silicon nanowire interface characteristics.[6,7]

Generally, the quality of the HfO_2/SiO_2 interface has a significant impact on device properties including threshold voltage (V_{th}), subthreshold slope (SS), and mobility. The transistor performance can be increased by optimizing the SiO_2 interface, resulting in improved current drivability, decreased power consumption, and enhanced device reliability. Overall, the *n*-type GAA-JL-NWFET-based oxygen gas sensor is a sensitive and effective way to detect and measure

Table II presents a sensitivity ($S_{I_{OFF}}$) comparison of the gate stack GAA-JL-NWFET with bulk MOSFET and GAA MOSFET using catalytic metal (Au) gate electrodes.

Shifting in Pd work function (mV)	$S_{I_{OFF}} = \frac{I_{OFF(after\ gas\ reaction)}}{I_{OFF(before\ gas\ reaction)}}$		
	Previously designed device[13]		Proposed device
	Bulk MOSFET $t_{Si} = 20$ nm	GAA MOSFET $t_{Si} = 20$ nm R = 10 nm	Gate stack GAA-JL-NWFET $t_{Si} = 10$ nm R = 5 nm
$\Delta\Phi_m = 50$	5.08	5.96	6.30
$\Delta\Phi_m = 100$	4.56	33.10	39.00
$\Delta\Phi_m = 150$	102	151	252.00

Device parameters: drain, source, and channel doping (N_{Si}) = 10^{19} cm^{-3}, oxide thickness (HfO_2 and SiO_2) is 1.5 and 0.3 nm, and their dielectric constants are 25.0 and 3.90, respectively at V_{GS} = 0.6 V and V_{DS} = 0.05 V, with channel length (L) = 40 nm.

the amount of oxygen gas. Its physical properties and design make it effective for use in high-performance gas sensing in several areas, including environmental monitoring, industrial safety, and healthcare advancements.

Conclusion

In this work, using Silvaco TCAD simulation analysis, we verified a gate stack junctionless GAA nanowire transistor with a gold (Au) metal gate as a feasible sensor for sensing oxygen gas based on an electrical detection method. The performance of the proposed device in terms of gas sensing provides better sensitivity than bulk MOSFET and GAA MOSFET in gas detection due to its optimized ultrasmall-scale surrounding gate and better surface-to-volume ratio. Operating a device in the subthreshold region is also desirable since it provides a low-cost and low-power gas sensor. This demonstrates that decreasing the silicon nanowire radius can increase the device sensitivity. The shift in subthreshold current produced by the variation in the work function of the gate metal due to the reactivity of gas molecules at the surface of the gate catalytic metal is utilized as the sensitivity parameter. This approach gives much greater sensitivity than threshold voltage variation or ON-current shift. In addition, long-term performance, consistency across sensors and batches, calibration reliability, and reaction to possible interferents are all evaluated as part of the stability and reproducibility assessments for the gate stack GAA-JL-NWFET oxygen gas sensors. These evaluations are necessary to guarantee the sensor's efficacy and dependability under varied operating circumstances. Table II shows that for a 100 meV gold metal gate electrode work function variation, the percentage improvement in subthreshold current sensitivity ($S_{I_{OFF}}$) is 15.13% and 88.31%, respectively, when the gate stack GAA-JL-NWFET is compared to the GAA MOSFET and bulk MOSFET devices. Overall, the *n*-type gate stack GAA-JL-NWFET-based oxygen gas sensor has higher sensitivity, lower power consumption, easier fabrication, better performance and miniaturization potential, and a wider operating temperature range than conventional oxygen sensors. These characteristics make it a promising candidate for developing highly sensitive and reliable oxygen gas detection applications in various fields and industries.

Acknowledgments Ethiopia's Ministry of Higher Education and Delhi Technological University's Microelectronics Research Laboratory are acknowledged for providing this study.

Conflict of interest The authors declare that there are no competing interests.

References

1. I. Sarangadharan, A.K. Pulikkathodi, C.-H. Chu, Y.-W. Chen, A. Regmi, P.-C. Chen, C.-P. Hsu, and Y.-L. Wang, Review—high field modulated FET biosensors for biomedical applications. *ECS J. Solid State Sci. Technol.* 7, Q3032 (2018).
2. S. Singh, M. Khosla, G. Wadhwa, and B. Raj, Design and analysis of double-gate junctionless vertical TFET for gas sensing applications. *Appl. Phys. A. Phys. A* 127, 1 (2021).
3. *Design and Performance Analysis of GAA Schottky Barrier-gate Stack Dopingless Nanowire Phosphine Gas Sensor.Pdf.*
4. Y. Pratap, M. Kumar, M. Gupta, S. Haldar, R.S. Gupta, and S.S. Deswal, *Sensitivity Investigation of Gate-All-around Junctionless Transistor for Hydrogen Gas Detection*, 2016 IEEE Int. *Nanoelectron. Conf.* 1, 1 (2016).
5. S. Mokkapati, N. Jaiswal, M. Gupta, and A. Kranti, Gate-All-around nanowire junctionless transistor-based hydrogen gas sensor. *IEEE Sens. J.* 19, 4758 (2019).
6. R. Chaujar and M. G. Yirak, *Sensitivity investigation of junctionless gate-all-around silicon nanowire field-effect transistor-based hydrogen gas sensor*, Silicon (2022).
7. A. Kumar, Palladium-based trench gate MOSFET for highly sensitive hydrogen gas sensor. *Mater. Sci. Semicond. Process.* 120, 105274 (2020).
8. G.B. Pour and L.F. Aval, Highly sensitive work function hydrogen gas sensor based on PdNPs/SiO_2/Si structure at room temperature. *Results Phys.* 7, 1993 (2017).
9. K. Tamersit and F. Djeffal, Double-Gate graphene nanoribbon field-effect transistor for DNA and gas sensing applications:

simulation study and sensitivity analysis. *IEEE Sens. J.* 16, 4180 (2016).
10. N. Van Toan, N. Viet Chien, N. Van Duy, H. Si Hong, H. Nguyen, N. Duc Hoa, and N. Van Hieu, Fabrication of highly sensitive and selective H_2 gas sensor based on SnO_2 thin film sensitized with microsized Pd Islands. *J. Hazard. Mater.* 301, 433 (2016).
11. J. Modi and B. H. Soni, *A Comparative Study of Sub-Threshold Swing for Different TFET Structures Performance of TFET Device*, **I**, 42 (2014).
12. J. Madan, R. Pandey, and R. Chaujar, Conducting polymer based gas sensor using PNIN-Gate all around-tunnel FET. *SILICON* 12(12), 2947 (2020).
13. R. Gautam, M. Saxena, R.S. Gupta, and M. Gupta, Gate-all-around nanowire MOSFET with catalytic metal gate for gas sensing applications. *IEEE Trans. Nanotechnol.* 12, 939 (2013).
14. C.H. Kim, I.T. Cho, J.M. Shin, K.B. Choi, J.K. Lee, and J.H. Lee, A new gas sensor based on MOSFET having a horizontal floating-gate. *IEEE Electron Device Lett.* 35, 265 (2014).
15. T. Usagawa and Y. Kikuchi, *A Pt-Ti-O Gate Si-Metal-Insulator-Semiconductor Field-Effect Transistor Hydrogen Gas Sensor*, J. Appl. Phys. **108**, (2010).
16. Y. Zhang, J. Zhang, B. Zhang, R. Si, B. Han, F. Hong, L. Sun, L. Li, B. Qiao, K. Sun, J. Huang, and M. Haruta, Boosting the catalysis of gold by O_2 activation at Au-SiO_2 interface. *Nat. Commun.* 1, 558 (2020).
17. N.F. Kosmani, F.A. Hamid, and M.A. Razali, Effects of high-k dielectric materials on electrical performance of double gate and gate-all-around MOSFET. *Int. J. Integr. Eng.* 12, 81 (2020).
18. G. Dhiman, Investigation of junction-less double gate MOSFET with high-k gate-oxide and metal gate layers. *Int. J. Innov. Res. Sci. Eng. Technol.* 8, 289 (2019).
19. Q. Li, H. Zhu, H. Yuan, O. Kirillov, D. Ioannou, J. Suehle, and C.A. Richter, A Study of metal gates on HfO_2 using Si nanowire field effect transistors as platform. *ECS Meet. Abstr.* Ma2012-02, 2614 (2012).
20. V. Mishra, L. Agarwal, C. Tiwari, and S. K. Gupta, *Potential Sensing Application of Oxygen Gas with Charge Plasma Based Negative Capacitance Tunnel FET*, Silicon (2022).
21. M. Gupta and V.P.-H. Hu, Influence of channel doping on junctionless and negative capacitance junctionless transistors. *ECS J. Solid State Sci. Technol.* 10, 065009 (2021).
22. J. Robertson, High dielectric constant oxides. *Eur. Phys. Journal-Applied Phys.* 184, 177 (2006).
23. U. Sharma, G. Kumar, S. Mishra, and R. Thomas, Advancement of gate oxides from SiO_2 to High-k dielectrics in microprocessor and memory. *J. Phys. Conf. Ser.* 2267, 012142 (2022).
24. G.D. Wilk, R.M. Wallace, and J.M. Anthony, High-κ gate dielectrics: current status and materials properties considerations. *J. Appl. Phys.* 89, 5243 (2001).
25. W.M. Sachtler, G.J.H. Dorgelo, and A.A. Holscher, The work function of gold. *Surf. Sci.. Sci.* 5, 221 (1966).
26. D.S. Software, ATLAS User's Manual. *IEEE Electron Device Lett.* 408, 567 (2013).
27. N. Gupta and R. Chaujar, Influence of Gate metal engineering on small-signal and noise behaviour of silicon nanowire MOSFET for low-noise amplifiers. *Appl. Phys. A Mater. Sci. Process.* 122, 1 (2016).
28. Z.X. Chen, H.Y. Yu, N. Singh, N.S. Shen, R.D. Sayanthan, G.Q. Lo, and D. Kwong, Demonstration of tunneling FETs based on highly scalable vertical silicon nanowires. *IEEE Electron Device Lett.* 30, 754 (2009).
29. A. Goel, S. Rewari, S. Verma, and R. S. Gupta, *Dielectric Modulated Triple Metal Gate All Around MOSFET (TMGAA)for DNA Bio-Molecule Detection*, in *2018 IEEE Electron Devices Kolkata Conference (EDKCON)*, Vol. 1 (IEEE, 2019), pp. 337–340.
30. S. P. Pandey, R. S. Kushwah, S. B. Singh, and S. Akashe, *Analytical Modeling and Comparison of Triple Gate MOSFET with Double Gate MOSFET*, 2013 Int. Conf. Control. Comput. Commun. Mater. ICCCCM 2013 (2013).
31. C.W. Hung, K.W. Lin, R.C. Liu, Y.Y. Tsai, P.H. Lai, S.I. Fu, T.P. Chen, H.I. Chen, and W.C. Liu, On the hydrogen sensing properties of a Pd/GaAs transistor-type gas sensor in a nitrogen ambiance. *Sens. Actuators B Chem.* 125, 22 (2007).
32. J. Madan, R. Pandey, and R. Chaujar, Conducting polymer based gas sensor using PNIN- gate all around - tunnel FET. *Silicon* 12, 2947 (2020).
33. N. Gupta and R. Chaujar, Optimization of high-k and gate metal workfunction for improved analog and intermodulation performance of gate stack (GS)-GEWE-SiNW MOSFET. *Superlattices Microstruct.* 97, 630 (2016).
34. E. Gnani, S. Reggiani, M. Rudan, and G. Baccarani, Effects of high-κ: (HfO_2) gate dielectrics in double-gate and cylindrical-nanowire FETs scaled to the ultimate technology nodes. *IEEE Trans. Nanotechnol.* 6, 90 (2007).
35. F. Garcia Ruiz, I.M. Tienda-Luna, A. Godoy, L. Donetti, and F. Gamiz, Equivalent oxide thickness of trigate SOI MOSFETs with high-κ insulators. *IEEE Trans. Electron Devices. Electron. Devices* 56, 2711 (2009).
36. S.L. Tripathi, R. Mishra, and R.A. Mishra, Multi-gate mosfet structures with high-K dielectric materials. *J. Electron. Devices* 16, 1388 (2012).
37. H. Ilatikhameneh, S. Kim, K. Ng, and G. Klimeck, *Optimum High-k Oxide for the Best Performance of Ultra-Scaled Double-Gate MOSFETs*, (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.

World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

Investigation of wire EDM process parameter on kerf width for high chromium high carbon steel by response surface methodology

Daniel Gnanaprakasam [1,*], Girish kumar [2] and Suhail Ahmad Siddiqui [1]

[1] *Department of Mechanical Engineering, Al-Falah University, Dhauj, Faridabad, India.*
[2] *Department of Mechanical Engineering, Delhi Technological University, Delhi, India.*

World Journal of Advanced Engineering Technology and Sciences, 2023, 08(01), 062–068

Publication history: Received on 01 December 2022; revised on 09 January 2023; accepted on 11 January 2023

Article DOI: https://doi.org/10.30574/wjaets.2023.8.1.0015

Abstract

The current work focused on studying the effect of process parameter on kerf width in Wire EDM using molybdenum wire electrode. Pulse on time, pulse off time, Servo voltage and wire feed were selected as a process parameter for design the DoE in Response Surface Methodology. It was observed that electrical parameters chosen had a significant impact on the Kerf Width. Pulse on time is the most important electrical characteristic for Kerf Width. Higher duration of the spark causes the job to shift dimensions. Utilizing Design Expert's point prediction functionality, model validation was carried out.

Keywords: Kerf width; Response Surface Methodology; Wire EDM; Design expert

1. Introduction

Wire EDM is one of the unconventional machining processes, which is mainly used to cutting of hard material like haste alloy, Inconel etc. In WEDM there is no direct contact between the wire and the workpiece. The machining behaviour is purely depends on electrical characteristics and not related to mechanical properties. The important electrical properties which influence the cutting rate is peak current, pulse on time, pulse off time, servo voltage. By varying the parameter in a certain range would certainly change the cutting rate and other outcome responses. Ramesh Kumar et.al machined the D2 steel material by considering the few important parameters like peak current, pulse on and off time. The kerf width was the assessment parameter on WEDM. Finally they concluded that the peak current has most impact on the kerf width when correlate with other parameter. [2]. Udhya Prakash et al performed machining on aluminium metal matrix composite to understand the behaviour of WEDM with the selected control variables. The MRR and SR was selected as output responses. The optimum solution was derived for the selected material with the range [3]. YasuhiroOkamoto et al observed the functionality of the multi wire EDM during the slicing of the semiconductor material. During the machining they considered kerf with as an assessment parameter. The wire feed rate effect on kerf width was finally reported in this study [4]. P. Saramah and P.K. Patawori reported the performance of WEDM on machining of Al based metal matrix composite. The analysis was done on Kerf width and MRR with the selected material. The considered parameter was pulse on and off time, Peak current. Pulse on have major influence on kerf width over others [5]. R.Calvo and M.Daniel conducted study on machining of high speed steel (HSS) AISI M42 in WEDM. The main focus was given on assessing the surface roughness of the selected steel material. The Taguchi approach was used to conduct the experiment. ANNOVA was done to predict the most influencing parameter [6]. P.Aswin and R.G Mote analysed the errors in the edge when making sharp corner In WEDM. The finite element analysis technique was employed to predict thermal effect in tip deflection [7]. R.Vijaya kumar et al preferred Response surface methodology approach to machine super alloy Inconel 625 in WEDM. The trail was done using CCD design. The Erosion rate and SR was studied with predominate factors. The different diameter sizes of wires been opted for study. The optimum result was derived through desirability analysis [8]. Ahmed A. A. Alduroobi et.al conducted experiment on AISI 1045 medium

* Corresponding author: G Daniel

Copyright © 2023 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

carbon steel to optimize the control variable in WEDM. The significant process parameter was selected and ANN technique was utilized to optimize the parameter on MRR and SR [9]. Sanjeev Kumar Garg observed the performance of WEDM during the machining of Al based MMC material. The design was effectively planned using Response surface methodology approach. The MRR and SR was Evaluated and ANNOVA was done for analysis. Genetic algorithm programme was generated using MATLAB and optimum result was derived [10]. Yasir Nawaz preferred WEDM to machine recently developed DC53 die steel. During the machining they observed MRR, SR and Kerf width with selected parameter. Response surface methodology was opted to conduct the experimental procedure [11]. M. Altin Karatas and M.A. Biberci developed a ANN model kerf width and SR for Ti-6Al-4V alloy in WEDM. Fluid pressure, servo voltage and feed rate were opted as control variable. ANNOVA test also conducted to conclude the influencing parameter [12]. Luca Watanabe Reolon et.al evaluated the performance of WEDM (kerf width & MRR) on Inconel 718 by using coated and uncoated zinc wire. The wire run off speed, pulse interval time and feed rate was considered as parameter [13]. K. Mandal et.al investigated the machinability of Al 6061 alloy using taguchi design philosophy. The MRR, Kerf width and SR was analysed. The surface properties were studied by SEM images [14]. From the extensive literature study, It found that there a need of analysing the High chromium high Carbon Die steel in die making industry. Further, the effect of process parameter like peak current , pulse on time, pulse off time and servo voltage on kerf width during the HCHCR steel is not been reported. So here an attempt is initiated to study the above said parameter on HCHCR steel using response surface methodology in WEDM is reported.

2. Material and methods

In this current study HCHCR steel is selected to machine in ELEKTRA SPRINTCUT 734 wire EDM machine. HCHCR steel is preferred in making of Dies for various applications. The specific properties of this material made to prefer the material in the particular domain. The application found in Thread rolling dies, Reaming tool, finishing roller for tire mills. Table 1 shows the chemical composition of HCHCR. It has been observed from the literature survey that most of WEDM studies were carried out using only a brass wire electrode. But in this study, experiments have been performed using a pure molybdenum wire as shown in Figure 1. The size of wire was measured before and after every trial experiment using a micrometre. It was observed that 0.01 mm of wire diameter is getting reduced after one hour continuous cutting operations which were conducted using fixed machining parameters.

2.1. Response surface Methodology

RSM (response surface methodology) is a statistical and trial technique. RSM requires a sufficient number of trial procedures information to deconstruct the problem and create numerical models for some information parameters and yield responses [10]. The needy factor is viewed as a surface to which a scientific sculpts is fitted. Four control factors are set at three levels in the current study. Along these lines, a standard central composite design (CCD) for demonstrating the Wire EDM factors for kerf width has been received. This outline includes a full factorial with 30 runs. The developed experimental DoE is shown in Table 2. Design expert 7 is used to generate the model for Kerf width. The exploratory data was subjected to an analysis of variance (ANOVA) to determine the model's behavior.

Table 1 Chemical composition of High Carbon High Chromium Die Steel (HCHCR) (wt %)

C	Cr	Si	V	Mn	Mo
1.54	12	0.32	0.91	0.34	0.76



Figure 1 Molybdenum Wire Electrode

Table 2 Input variable and their levels

Variables	**Symbol**	**Levels**				
		-2	**-1**	**0**	**1**	**2**
Pulse on Time- Ton (µs)	A	105	110	115	120	125
Pulse off time T-off (µs)	B	35	40	45	55	10
Servo Voltage (volt)- Sv	C	10	15	20	25	30
Wire feed rate (m/min)- Wf	D	4	6	8	10	12

2.2. Experimental Set Up

On the four-axis ELEKTRA SPRINTCUT 734 machine, analyses were conducted to consider how control variables at different stages affect kerf the width. The sparks are produced among the work piece as well as the wire terminal. Dielectric liquid is continuously supplied between the wire electrode and work surface. The work setup is shown in the Figure 1. The minute particle generated during the machining was flushed by the dielectric medium. The size of the work plate is 90mm x 75mm x 8mm. The wire of 0.25mm diameter is used for the experimentation. The demineralized water at room temperature is used as a dielectric liquid. Kerf width is the summation of wire diameter and two times of wire workpiece gap [13,15].

$$\text{Kerf width (Kw)} = 2 \times \text{gap between wire and workpiece}$$

Table 3 Design of experiments

Run	Pulse on time	Pulse off time	Servo voltage	Wire feed	Kerf Width
1	2.00	0.00	0.00	0.00	0.191
2	1.00	1.00	-1.00	1.00	0.182
3	-1.00	1.00	1.00	-1.00	0.141
4	-1.00	1.00	-1.00	-1.00	0.139
5	-1.00	-1.00	1.00	-1.00	0.139
6	-2.00	0.00	0.00	0.00	0.132
7	1.00	1.00	1.00	-1.00	0.182
8	1.00	1.00	-1.00	-1.00	0.179
9	1.00	-1.00	-1.00	-1.00	0.175
10	1.00	1.00	1.00	1.00	0.175
11	0.00	0.00	0.00	-2.00	0.159
12	0.00	0.00	0.00	2.00	0.155
13	0.00	0.00	0.00	0.00	0.154
14	0.00	0.00	0.00	0.00	0.154
15	0.00	0.00	0.00	0.00	0.154
16	-1.00	1.00	-1.00	1.00	0.152
17	-1.00	1.00	1.00	1.00	0.151
18	-1.00	-1.00	1.00	1.00	0.151
19	-1.00	-1.00	-1.00	1.00	0.149
20	-1.00	-1.00	-1.00	-1.00	0.145
21	0.00	0.00	0.00	0.00	0.164
22	0.00	0.00	2.00	0.00	0.161
23	0.00	0.00	0.00	0.00	0.161
24	0.00	0.00	0.00	0.00	0.161
25	0.00	2.00	0.00	0.00	0.161
26	1.00	-1.00	-1.00	1.00	0.172
27	1.00	-1.00	1.00	-1.00	0.169
28	1.00	-1.00	1.00	1.00	0.169
29	0.00	-2.00	0.00	0.00	0.166
30	0.00	-2.00	0.00	0.00	0.165

3. Results and discussion

3.1. Effect of WEDM parameters on kerf width

Response surface charts were plotted to investigate the effect of Wire EDM parameters on kerf width, as shown in the figure. 2 & 3. The RSM chart shows that the KW increases with increasing pulse on time and Wire Feed. Increasing the

pulse on time increases the discharge energy across the electrode, resulting in increased melting and evaporation of material [12]. Increased melting and evaporation, combined with a high dielectric supply, results in a wider kerf. [12,13].

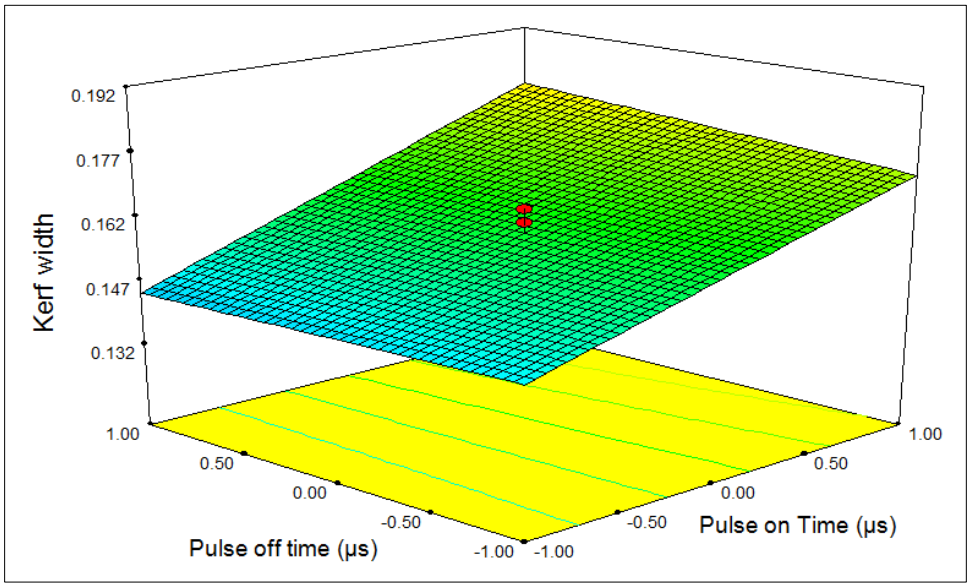


Figure 2 KW vs Pulse on time & Pulse off time

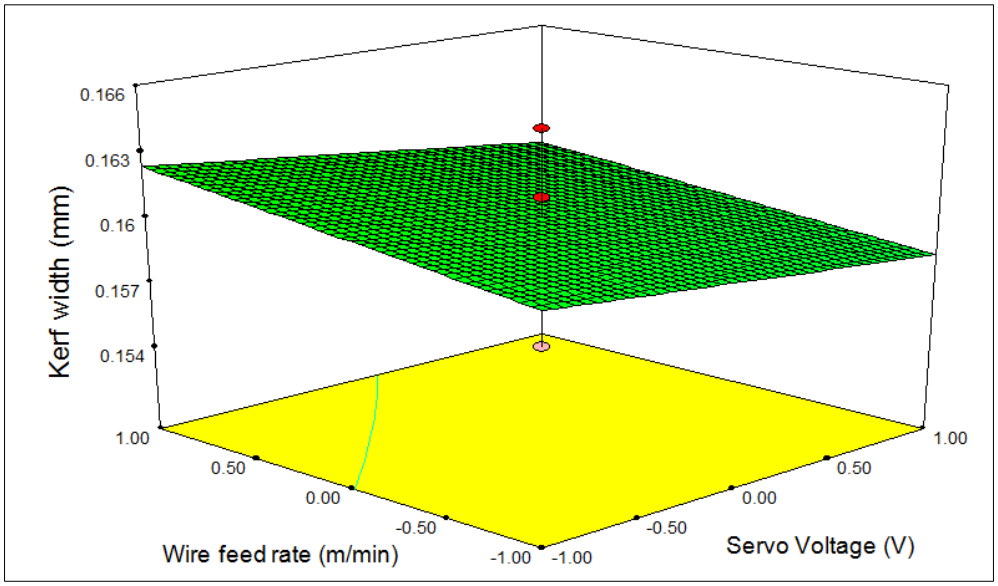


Figure 3 KW vs servo voltage & Wire feed rate

3.2. Analysis of kerf width

Table 4 shows that the quadratic model's p value has a significant effect on the output response. As confirmed by ANOVA results, it is understandable that all input parameters have a significant impact on kerf width. At the same time, the kerf width increased by a percentage, when pulse on time (Ton) increased from low to high. Long pulse on time (Ton) leads to high performance in heat transfer inside the dielectric fluid and work piece, and it is capable of removing molten or liquid metal by making dielectric pressure flushing in a constant certain amount. F ratio was greater than the tabulated values at 95 % confidence level. R^2 and adjusted R^2 values are 91.70% & 88.54% respectively. The developed regression equation is shown in the equation 1.

KW =+0.16028+0.014754 * A +9.20833E-004* B-1.01250E-003 * C +1.01250E-003 * D
+2.11875E-003 * A * B-6.18750E-004* A * C+6.31250E-004* B * C+4.06250E-00 * B * D

Table 4 ANOVA for Ker width

	Sum of		**Mean**	**F**	**p-value**	
Source	**Squares**	**df**	**Square**	**Value**	**Prob > F**	
Model	0.005381	8	0.000673	29.1600428	< 0.0001	significant
A-A	0.0052245	1	0.005224	226.494528	< 0.0001	
B-B	2.035E-05	1	2.04E-05	0.88224744	0.3583	
C-C	2.46E-05	1	2.46E-05	1.06664133	0.3135	
D-D	2.46E-05	1	2.46E-05	1.06664133	0.3135	
AB	7.183E-05	1	7.18E-05	3.1138416	0.0922	
AC	6.126E-06	1	6.13E-06	0.26556297	0.6117	
BC	6.376E-06	1	6.38E-06	0.27640116	0.6046	
BD	2.641E-06	1	2.64E-06	0.11447847	0.7385	
Residual	0.0004844	21	2.31E-05			
Lack of Fit	0.0003812	16	2.38E-05	1.15418702	0.4754	not significant
Pure Error	0.0001032	5	2.06E-05			
Cor Total	0.0058654	29				

4. Conclusion

In this study to explore the kerf width on machining of High carbon and high chromium die steel were done by using Response surface methodology design. Based on these experiments the following conclusions are drawn.

- Wire feed rate and pulse on time (Ton) have significant impact on and kerf width (Kw). Kerf width is increased with increase in wire feed rate and Ton, and decreased on increase in pulse off time. .
- (ii)The maximum kerf width is observed on 125 µs pulse on time, 45 µs pulse off time, 20V voltage and, 8m/min wire feed.
- (iii)The least kerf width is observed on 105 µs pulse on time, 45 µs pulse off time, 20V voltage and, 8m/min wire feed.
- (iv)ANOVA results show that model is significant, lack of fit is not significant and pulse on time has a significant effect on kerf width rather than a other parameter

Compliance with ethical standards

Acknowledgments

I acknowledge my gratefulness to Dr. K.Sivakumar, Dean Bannari Amman Institute of Technology (BIT), Sathyamangalam, Erode District in Tamil Nadu and Dr.P.Mathan Kumar, Assistant professor, Department of Mechanical Engineering, Bannari Amman Institute of Technology(BIT), Sathyamangalam, Erode District in Tamil Nadu for allowing me to utilize the lab facility to carry out the experiments.

Disclosure of conflict of interest

The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report. We certify that the submission is original work and is not under review at any other publication.

References

[1] Kumar, N., Kumari, S., Abhishek, K., Nandi, G., & Ghosh, N. (2022). Study on various parameters of WEDM using different optimization techniques: A review. Materials Today: Proceedings, 62, 4018–4024. https://doi.org/10.1016/j.matpr.2022.04.596

[2] Verma, R. K., Upadhyay, R. K., & Rizvi, S. A. H. (2022). Modeling of kerf width during WEDM of die steel D3. Materials Today: Proceedings, 54, 832–838. https://doi.org/10.1016/j.matpr.2021.11.115

[3] Udaya Prakash, J., Sivaprakasam, P., Jebarose Juliyana, S., Ananth, S., Sarala Rubi, C., & Divya Sadhana, A. (2022). Multi-objective optimization using grey relational analysis for wire EDM of aluminium matrix composites. Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2022.09.415

[4] Okamoto, Y., Ikeda, T., Kurihara, H., Okada, A., & Kido, M. (2018). Control of Kerf Width in Multi-wire EDM Slicing of Semiconductors with Circular Section. Procedia CIRP, 68, 100–103. https://doi.org/10.1016/j.procir.2017.12.030

[5] Sarmah, P., & Patowari, P. K. (2022). Machinability study of Al 6063-based MMCs with SiC reinforcement particles using WEDM process. Materials and Manufacturing Processes, 1–14. https://doi.org/10.1080/10426914.2022.2149786

[6] Calvo, R., & Daniel, M. (2019). Wire electrical discharge machining (EDM) setup parameters influence in functional surface roughness. Procedia Manufacturing, 41, 602–609. https://doi.org/10.1016/j.promfg.2019.09.048

[7] Aswin, P., & Mote, R. G. (2022). Numerical study of transient thermal profile and thermally deformed tip shape in Wire EDM of narrow angles. Journal of Manufacturing Processes, 81, 638–654. https://doi.org/10.1016/j.jmapro.2022.07.021

[8] Vijayakumar, R., Srirangarajalu, N., Santhanakumar, M., & Adalarasan, R. (2022). Investigation in μ-WEDM of Inconel 625 superalloy using RSM-CCD technique. Materials and Manufacturing Processes, 1–12. https://doi.org/10.1080/10426914.2022.2116035

[9] Alduroobi, A. A. A., Ubaid, A. M., Tawfiq, M. A., & Elias, R. R. (2020). Wire EDM process optimization for machining AISI 1045 steel by use of Taguchi method, artificial neural network and analysis of variances. International Journal of System Assurance Engineering and Management, 11(6), 1314–1338. https://doi.org/10.1007/s13198-020-00990-z

[10] Garg, S.K., Manna, A. & Jain, A. Investigation and Analysis of Surface Veracity and Parametric Aspects During WireEDM of Al/ZrO2(p)-Metal Matrix Composite. Arab J Sci Eng 47, 8417–8438 (2022). https://doi.org/10.1007/s13369-021-05531-9

[11] Nawaz, Y., Maqsood, S., Naeem, K., Nawaz, R., Omair, M., & Habib, T. (2020). Parametric optimization of material removal rate, surface roughness, and kerf width in high-speed wire electric discharge machining (HS-WEDM) of DC53 die steel. The International Journal of Advanced Manufacturing Technology, 107(7–8), 3231–3245. https://doi.org/10.1007/s00170-020-05175-3

[12] Altin Karataş, M., & Biberci, M. (2022). Statistical Analysis of WEDM Machining Parameters of Ti-6Al-4V Alloy Using Taguchi-Based Grey Relational Analysis and Artificial Neural Network. Experimental Techniques. https://doi.org/10.1007/s40799-022-00601-5

[13] Reolon, L. W., Henning Laurindo, C. A., Torres, R. D., & Amorim, F. L. (2018). WEDM performance and surface integrity of Inconel alloy IN718 with coated and uncoated wires. The International Journal of Advanced Manufacturing Technology, 100(5–8), 1981–1991. https://doi.org/10.1007/s00170-018-2828-6

[14] Mandal, K., Sarkar, S., Mitra, S., & Bose, D. (2018). Analysis of Wire-EDM Input Parameters on Kerf Width and Surface Integrity for Al 6061 Alloy. Innovation in Materials Science and Engineering, 35–41. https://doi.org/10.1007/978-981-13-2944-9_5

[15] Reolon, L. W., Henning Laurindo, C. A., Torres, R. D., & Amorim, F. L. (2018). WEDM performance and surface integrity of Inconel alloy IN718 with coated and uncoated wires. The International Journal of Advanced Manufacturing Technology, 100(5–8), 1981–1991. https://doi.org/10.1007/s00170-018-2828-6

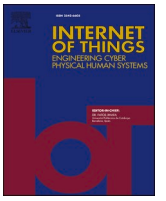
Internet of Things 25 (2024) 101098



Contents lists available at ScienceDirect

Internet of Things

journal homepage: www.sciencedirect.com/journal/internet-of-things



Review article

IoT-driven augmented reality and virtual reality systems in neurological sciences

Mehar Sahu [a], Rohan Gupta [a,b], Rashmi K. Ambasta [a,c], Pravir Kumar [a,*,1]

[a] Molecular Neuroscience and Functional Genomic Laboratory, Department of Biotechnology, Delhi Technological University, Delhi, India
[b] School of Medicine, University of South Carolina, Columbia, SC, United States of America
[c] Department of Biotechnology and Microbiology, SRM University, Sonipat, Haryana, India

ARTICLE INFO

Keywords:
Augmented reality
Brain–computer interface
Neuromodulation
Neuroprosthetic
Neurosurgery
Virtual reality

ABSTRACT

Research in augmented and virtual reality in congregation with the Internet of Things has opened many avenues in diagnosing and treating neurological disorders. Augmented reality permits inserting virtual content in the real world, while virtual reality is a simulated experience that provides an artificial three-dimensional environment to the user. These are the game-changer technologies as they give a transformational change to existing technologies and methods. Augmented and virtual reality has come out as a significant technology in treating various mental disorders, thereby providing great applications in the field of neuroscience. In this review, we shed light on different components required for developing an augmented and virtual reality-

Abbreviations: AR, augmented reality; VR, virtual reality; HMD, head-mount display; CAVE, cave automatic virtual environment; 3D, three-dimensional; SNVs, single nucleotide variations; GWAS, genome-wide association studies; MDV, molecular dynamics visualization; HIVE, hub for immersive visualization and e-research; iMD-VR, interactive molecular dynamics in virtual reality; ARM, augmented reality microscope; AI, artificial intelligence; GUI, graphical user interface; CT, computed tomography; MRI, magnetic resonance imaging; EEVIS, enhanced electrophysiology visualization and interaction system; MR, mixed reality; GAINS, goggle augmented imaging and navigation system; EEG, electroencephalography; SLAM, simultaneous localization and mapping; GNSS, global navigation satellite system; LiDAR, light detection and ranging; RGB, red, green, blue; EKF, extended Kalman features; BCI, brain computer interface; ECoG, electrocorticogram; fMRI, functional magnetic resonance imaging; fNIRS, functional infrared spectroscopy; MEG, magnetoencephalography; PET, positron emission tomography; SMAII, spinal muscular atrophy type II; SSVEP, steady-state visual evoked potentials; ERD/ERS, (de)synchronization; ALS, amyotrophic lateral sclerosis; LED, light-emitting diode; RGS, rehabilitation gaming system; LCD, liquid-crystal display; HSV, hue, saturation, and value; TBI, traumatic brain injury; PASAT, paced auditory serial addition test; EMG, electromyography; MUPs, motor unit potential; CNN, convolutional neural network; EMG-BF, EMG biofeedback; NDDs, neurodegenerative diseases; REM, rapid eye movement; PD, Parkinson's disease; MEP, motor-evoked potential; TMS, transcranial magnetic stimulation; CST, corticospinal tract; ML, machine learning; DL, deep learning; AD, Alzheimer's disease; BEAPP, batch EEG automated processing platform; BIDS, brain imaging data structure; HAPPE, Harvard automated processing pipeline for EEG; WSBDS, wire-embedded silicon-based dry-contact sensor; WSN, wireless sensing network; IoT, Internet of Things; BS, base station; MCPS, medical cyber-physical systems; MS, multiple sclerosis; MCI, mild cognitive impairment; VMET, VR version of multiple errand test; VEPF, visual evoked potentials and fields; RNN, recurrent neural network; NLP, natural language processing; QR, quick response; RFID, radio-frequency identification; SAD, social anxiety disorders; ARET, AR exposure therapy; VRET, VR exposure therapy; LMICs, low- and middle-income countries; IGNS, image-guided neurosurgery; MR, magnetic resonance; IBIS, intraoperative brain imaging system; CTA, CT angiography; MRA, magnetic resonance angiography; DSA, digital subtraction angiography; AVM, arteriovenous malformation; VIPAR, virtual interactive presence and AR; OST-HMD, optical see-through head-mounted display; RGB-D, RGB-depth; VEGS, virtual environment grocery store; FOV, field of view; OST, optical see-through; VST, virtual studio technology.

* Corresponding author at: Professor of Biotechnology and Dean International Affairs, Delhi Technological University (Formerly Delhi College of Engineering), Molecular Neuroscience and Functional Genomics Laboratory, Shahbad Daulatpur, Bawana Road, Delhi 110042, India.
E-mail address: pravirkumar@dtu.ac.in (P. Kumar).
[1] Scopus ID: 14831447800; ISI Research ID: B-2164-2015.

https://doi.org/10.1016/j.iot.2024.101098

Available online 2 February 2024
2542-6605/© 2024 Elsevier B.V. All rights reserved.

based technology followed by different devices used in the augmented and virtual-reality-based systems. We highlighted the use of this technology in the diagnosis of neurological defects and proposed strategies to use these advancements to provide an immersive experience for both research and educational purposes. Moreover, we show an extensive implementation of augmented and virtual reality in neurological surgery, neuromodulation, and neuroprosthetics. Therefore, the role of the Internet of Things with augmented and virtual reality for diagnosing and treating neurological disorders is the future.

1. Introduction

Artificial Intelligence (AI) has come up in producing intelligent machines and doing human-centric tasks that it is nearly touching all the fields. AI first emerged in 1958 when terms such as Machine Learning (ML) and Natural Language Processing (NLP) were developed. In 1970, rule-based approaches became an attraction. Around 2012, Deep Learning (DL) revolutionized the success of AI and presently we are experiencing tremendous growth in Virtual Reality (VR), Augmented Reality (AR), Mixed Reality (MR), and Extended Reality (XR) [1]. AR and VR provide a great avenue to the education and research community in an attempt to enrich the medical diagnosis and treatment experience with an attractive virtual and immersive environment [2]. AR is an emerging technology that allows inserting virtual content in the real world, whereas, VR is a simulated experience that provides an artificial three-dimensional (3D) environment, to the user, using different electronic components. MR involves both AR and VR, whereas, XR is an umbrella term that includes all AR, VR, and MR. However, XR and AI stand different from each other as XR is more focused on enhancing human experiences but AI attempts to replicate human ways [1].

Implementation of AR and VR technology in healthcare and medical systems has revolutionized medical training, surgical procedures, diagnostic purposes, physical fitness, and patient information [3]. Further, VR and AR technology has flourished as useful tool for medical diagnosis and imaging in the field of neurobiology and neuroscience. For instance, AR and VR have been extensively involved in cognitive research, evaluation, rehabilitation, gait abnormalities, surgical procedures, neuropathic pain management, and mapping brain activity [4]. These technologies prove to be efficient for helping a patient who has lost physical movement and cognitive disabilities. Further, different algorithms and tools have been designed for mapping brain activity based on AR and VR environments, such as brain–computer interface (BCI), simultaneous localization and mapping (SLAM), electroencephalography (EEG), electromyography (EMG), wireless sensing network (WSN), and rehabilitation gaming system (RGS). Apart from mapping brain activity, AR and VR systems have been extensively applied in neurosurgical procedures and training, namely neuro-oncology surgery, neurovascular surgery, and spinal surgery. Likewise, AR and VR devices and systems have been used for neuromodulation and neuroprosthetics, such as neuronavigation, episodic memory assessment, neurogenesis induction and neuroplasticity, and neuroanatomy [5]. Our review aimed at better understanding advancements in AI and ML algorithms with the Internet of Things (IoT) in the field of AR and VR for neuroscience research.

2. Algorithms/devices for mapping brain functions

2.1. Simultaneous localization and mapping

For many years, autonomous navigation has become an active research area for researchers worldwide, which requires high precision and accuracy. The two prerequisites were achieved through accurate information of the surrounding environment and the excellent localization ability of the robot itself. Today the most applicable navigation system used is the Global Navigation Satellite System (GNSS). However, GNSS ends up with some errors due to its dependence on the surrounding environment, which is not acceptable for autonomous navigation [6]. The only solution for the eradication of GNSS error is to induce the prior information of the surroundings. Thus, creating 3D models or maps is considered as the primary step in complete autonomous navigation. SLAM is an approach capable of mapping 3D environments accurately with intelligible genomic properties and complex semantic objects [7]. SLAM is used to map the 3D image of the surrounding environment using different sensors while predicting its own position in the surrounding environment, which is crucial for the autonomous navigation system [8]. SLAM approach can perform loop closure, which makes it advantageous over other navigating systems. The implementation of loop closure reduces the ambiguities of every pose of the map as every position in the loop appears to be correlated in the environment [9].

$$z_k = h(X_k, M_k) + e_k$$

SLAM is a prediction or estimation problem. Thus, we have to calculate both the X and M variables, which denote the trajectory of the robot and the position of the landmarks in the environment, respectively. In the above equation, e_k is a random measurement noise. With the advancement in technology and the growing need for autonomous navigation systems, there are different variations to traditional SLAM systems such as visual-SLAM, light detection and ranging (LiDAR)-SLAM, and LiDAR-camera fusion, which are further divided into different types based on the type of environment, features, and navigation parameters [10]. The application between SLAM and nature is not an emerging concept, and the researchers have demonstrated different brain-inspired SLAM models. NeuroSLAM and RatSLAM are two prominent brain-inspired SLAM models. NeuroSLAM is a novel neuro-inspire mapping system with four degrees of freedom in a 3D environment, where 3D grid cells and the multi-layered head direction cells denote the robot's four

degrees of freedom pose in a 3D environment [11]. Similarly, RatSLAM integrates udometric information with landmark information to create accurate 3D dimensions of the environment. RatSLAM can reduce both major and minor path integration errors and operate with obstruse landmark data [12]. Further, Chancan et al. constructed a neuroscience-based dynamical network model for visual plane recognition and navigation systems. The proposed model is based on temporal properties underlying spatial navigation in the brain. The results demonstrated that the proposed model yields a high-performance value compared to SeqSLAM [13]. Likewise, Zhou et al. proposed a brain-inspired SLAM system, which incorporated different oriented features and elementary features from vast accelerated segment tests and rotated binary robust test, respectively. The proposed method is an extended version of RatSLAM, which extracts features from red, blue, green (RGB) images and suitable to recognize the sequences of familiar visual scenes [14]. Another SLAM-based model was constructed by Bresson et al. to solve memory and processing time problems known as MSLAM, which is based on an extended Kalman filter (EKF). The advantage of MSLAM is that it catches incorrect updates and corrects them, and minimizes the effect of the linearization on the entire system [15]. Recently, Munoz-Montoya et al. employed AR technology for the assessment of

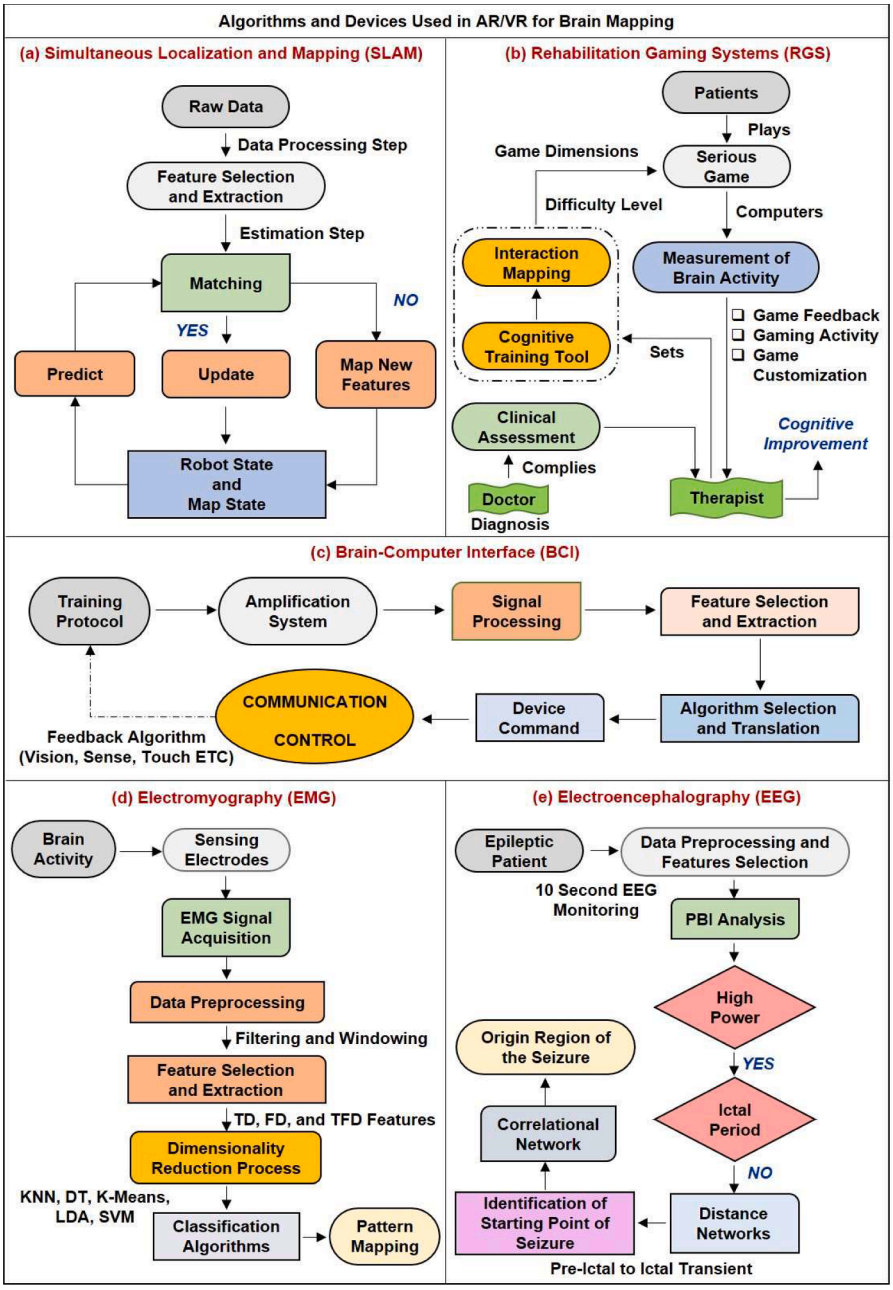


Fig. 1. Illustration of varied algorithms and devices for brain mapping acquiring AR/VR: (a) Simultaneous localization and mapping (SLAM); (b) Rehabilitation gaming systems (RGS); (c) Brain–computer interface (BCI); (d) Electromyography (EMG); and (e) Electroencephalography (EEG).

short-term spatial memory, which was separated into two phases, namely a learning phase and an evaluation phase. The performance evaluation of the proposed model demonstrated that perceived competence is directly proportional to the number of correctly placed shapes, and the number of attempts is inversely proportional to perceived competence. Besides, the results concluded that gender does not affect types of memory tasks [16]. Altogether, the evidence discussed above concluded that the performance of SLAM is higher than GNSS algorithm and different variations of SLAM have been constructed to increase accuracy and minimize memory consumption for efficient mapping and navigation. For instance, (Fig. 1(a)) represents how the raw data is processed which goes for feature selection and extraction, and further undergoes an estimation step for matching data. Afterward, an update of data is done which finally goes for robot state and map state. If matching is not properly done, mapping new features will go for the robot state and map state. The robot state passes information for prediction, which again follows the matching state. A neuroscience-based SLAM model to create cognitive maps has been proposed, which yields high performance and accuracy compared to traditional SLAM models.

2.2. Brain–computer interface

The advent of computer technology and massive research on the human brain during the past two decades led to exciting and fascinating development in BCI, making it a high-priority research area in applied science. The emergence of BCI paves the way for neurorehabilitation for patients with motor function impairment, stroke, brain injuries, and physically disabled people [17–19]. BCI is a technology used to exchange information between the computer and brain, which is divided into two categories: invasive BCI (electrocorticogram (ECoG) and intracortical recordings) and non-invasive (EEG) BCI systems [20–22]. Further, a non-invasive system has advantages over an invasive system due to spatial resolution in the range of millimeters, higher frequency bandwidth, amplitude above 100 μV, and decreased sensitivity to movement defects [23,24]. The important characteristic feature of BCI is that the data generated is being processed in real-time through either direct connections (EEG) or indirect connections (blood oxygen measurements, functional magnetic resonance imaging (fMRI), functional infrared spectroscopy (fNIRS)) [25–28]. The training protocol starts with an amplification system followed by the processing of signals, which passes information for feature extraction and selection followed by selection and translational of the algorithm. Afterward, the command device communicates and controls the information, which through the feedback algorithm passes information again to the training protocol (Fig. 1(c)).

In 1998, the first invasive BCI device was implanted in a human by Philip Kennedy, whereas, the foremost BCI-based game called "BrainGate" was introduced in 2003 by John Donoghue. Further, in 2004 Matt Nagle was the first patient implanted with a BCI system, whereas, Leuthardt et al. demonstrated that ECoG has an accuracy of 73–100 % and is a valuable source for controlling signals of the BCI system [29,30]. In 2014, first-time information was transferred between two brains without the application of any motor or peripheral nervous system [31]. In 2016, Tim Hemmes was implanted with a BCI-based system, which allowed him to recover the tactile sensation [32]. Later on, again in 2016, a 24-year-old patient was implanted with a Utah microelectrode array and trained to use his cortical neuronal activity [33]. Similarly, Ganzer et al. performed a study, which demonstrated patients implanted with an intracortical recording array in the primary motor cortex after chronically paralyzed had enhanced touch events and motor functions [34].

Moreover, apart from non-invasive BCI systems include magnetoencephalography (MEG), positron emission tomography (PET), fMRI, and fNIRS. Non-invasive systems were developed to find an alternative control and communication method for patients through devices such as a wheelchair, robotic arm, and video display [35–38]. Millan et al. concluded that asynchronous EEG-based BCI allowed operating of the wheelchair for a long time due to shared control system between BCI and the wheelchair, whereas, Cincotti et al. concluded that patients with Spinal Muscular Atrophy Type II (SMA II) when underwent SMR-BCI training in a controlled environment and mapped home area have able to control domestic appliances [39,40]. Further, the application of BCI extended to the rehabilitation of patients that are unable to speak, known as spellers through p300, steady-state visual evoked potentials (SSVEP), and event-related (de)synchronization (ERD/ERS) [41–44]. Peter et al. demonstrated the application of SSVEP-BCI in shuffle speller typing interface that is applied to amyotrophic lateral sclerosis (ALS) patients, whereas, Hwang et al. applied a QWERT keyboard together with a flickering LED for the SSVEP-BCI [45,46].

Further, due to high cost, non-invasive BCI will remain limited to public use, and thus there is a growing need to develop BCI devices that are cost-effective and have high efficiency. Today, wearable headsets based on non-invasive BCI have been developed frequently by different companies such as Emotive Inc., Ant Neuro, Cognionics, Neurosky, OpenBCI, InteraXon, and CREmdical [47–52]. Mounting evidence suggests that the future trends and application of BCI systems are strongly related to cost-effectiveness and the development of intelligent algorithms for the analysis of big data. Further, there is a growing need to reduce the calibration time of the current BCI systems, which yields high performance of the devices. To sum it up, the BCI-based devices can be applied to training, education, entertainment, neuroimaging, and medical assistance with the focus on real-world applications, which carry out the BCI applications from laboratory to clinical products [53]. Currently, BCI systems are used to treat stroke, psychiatric disorders, spinal cord injuries, cerebral palsy, chronic peripheral neuropathies, and muscular dystrophies [54–57]. However, BCI systems come with limitations such as signal-acquisition hardware, BCI validation, and reliability, which should be removed to apply BCI systems at a clinical level.

2.3. Rehabilitation gaming system

RGS is a VR-based rehabilitation technique that activates the entire neuronal system to transfer direct simulation to motor areas affected by neuronal injury through non-invasive multi-modal simulation. As depicted in (Fig. 1(b)), where, patients pass information for a serious game, which measures brain activity through the therapist. Therapists process information through cognitive training

tools and interaction mapping, which sets difficulty levels and game dimensions. For diagnosis purposes, the doctors compile information for clinical assessment and pass information to therapists. RGS, a generic rehabilitative approach, aims to engage patients with upper extremity dysfunction after stroke in a specific training program. A typical RGS consists of a standard PC with a Linux operating system, a 19" or above the LCD monitor, a 3D graphic card, a video camera, and a pair of data gloves [58–60]. Further, during training, the movement of arms is tracked with a vision-based tracking system known as AnTS at an update frequency of 35 Hz. In the RGS system, the head of the object and color patches present on the elbow and wrists are tracked through the video camera present on the top of the monitor [61]. For color detection, RGB data is converted into hue, saturation and value (HSV) data, whereas, for object detection, Bayesian inference techniques are used, which are based on Hue value, velocity vector, and object size. Recent years witnessed the implementation of RGS in the rehabilitation of patients with neurological defects such as cerebral palsy, stroke, motor function impairment, and upper extremities dysfunction, which engages subjects into a physical exercise in a friendly environment [62–65]. For instance, Daoud et al. proposed three-game-based exercises and a computerized assessment method for rehabilitation of patients with upper-limb cerebral palsy [66], whereas, Ballester et al. conducted a randomized controlled trial with 14 cerebral palsy subjects to assess its usability, effectiveness, and acceptability. The results demonstrated patient's impairment level is similar to that of the RGS adaptation algorithm [63]. Similarly, Muñoz et al. developed an RGB-based platform known as KineActiv (Microsoft Kinect v2) to rehabilitate upper limbs, allowing users to assess and examine the session results from patient's health [67]. Recently, the RGS system was used to rehabilitate patients with traumatic brain injury (TBI). For example, Välimäki et al. conducted a controlled-trail with 90 Finnish-speaking adults (18–65 age group) divided into three groups that are a rehabilitation gaming group ($n = 29$), an entertainment gaming group ($n = 29$), and a control group ($n = 32$). The study assesses patient outcomes through different tests such as visuomotor tasks, Simon task, WAIS-IV digit span and paced auditory serial addition test (PASAT), depression, self-efficacy, and executive functions [68]. The growing popularity of RGS and the simultaneous failure of traditional therapies for neurorehabilitation have led to the development of different VR/AR-based applications. Altogether, it must be concluded that the RGS method is a useful and crucial system for neurorehabilitation for many reasons. The prominent one is adapting the patient to the system, easy to use, and friendly interface (Table 1).

2.4. Electromyography

EMG is commonly implemented in clinical neurophysiology to record the electrical impulses of muscles and nerves at different levels. It is applied to diagnose neuropathies and myopathies due to differences in EMG signals compared to healthy controls [76,77]. In EMG, brain activity is measured with sensing electrodes, which acquire the EMG signal followed by data preprocessing and feature selection. Afterward, dimensionality reductions are carried out through different process, which passes information to classify data through different algorithms. The classification algorithms used are KNN, K-means, and support vector machine, which further maps brain activity patterns (Fig. 1(d)). Apart from its high applicability in diagnosis, monitoring EMG signals is considered one of the complex tasks as it is highly subjective. In addition, monitoring neuropathies' EMG signals is relatively easy compared to myopathies' EMG signals [78]. Neuropathies EMG signals have fibrillation potential, sharp waves, high amplitude, and long duration motor unit potential (MUPs), whereas, myopathies EMG signals have low amplitude, short-duration MUPs, and fatigue waves. The application of EMG varies from upper-limb prostheses, electric power wheelchairs, and human-computer interactions [79–81]. Recent times witnessed AI-based algorithms such as neural networks, fuzzy logic, support vector machines, and regression models for monitoring EMG signals. For instance, Lei Zhang estimated the upper limb movement from EMG through a backpropagation neural network. In his experiment Zhang, processed the EMG signals such as amplification, denoising, biasing, and normalization, followed by extraction of signal features, namely time domain, frequency domain, and time-frequency domain. The results demonstrated that the STM32 microcontroller reduces the size of control equipment compared to traditional systems, which provided convenience for the patient [82].

Table 1
AR/VR-based RGS applications involved in the rehabilitation of patients with neurological defects.

Application name	Disease	Description/Outcomes	Reference
RehabFork	Upper limb stroke	RehabFork is a device created for training of upper-limb movement, which provides easiness for the user to perform eating tasks.	[69]
iVG4Rehab system	Cerebral palsy	The iVG4Rehab system is capable of dynamically adjusting game parameters in accordance with the abilities and therapeutic needs of its users.	[70]
ViTAS Gaming Suite	Stroke	ViTAS is a combination of virtual and traditional rehabilitation. The programme is used for clinical aspects for rehabilitation purposes. Mostly, ViTAS is used in the rehabilitation of patients with stroke.	[71]
Mii-vitaliSe	Multiple sclerosis	Mii-vitaliSe is constructed for the use of physiotherapist it is based on supervised Nintendo Wii familiarization sessions.	[72]
Pedbothome	Cerebral palsy	The program is used for the rehabilitation of children with cerebral palsy.	[73]
GAMEREHAB@HOME	Neurological defects	It is a multisensory device for the rehabilitation of patients with joints problem. It is associated with camera and sensors, which records 3D movement of joints during clinical monitoring.	[74]
ReHabgame	Neurological defects	It is a Non-immersive VR RGS developed using Microsoft Kinect™ and the Thalmic™ Labs Myo gesture control armband. Psychometric characteristics were assessed for the performance evaluation of the subjects.	[75]
Move-IT	Upper limb stroke	The game integrates the oculus rift head-mounted display (HMD) and the leap motion hand tracker. The outcome of the study concluded that the subjects were pleased with the application of the proposed system.	[68]

Similarly, Ulkir et al. classified EMG signals through fuzzy logic, where gripping movements have been classified using features such as root mean square, wavelength length, and kurtosis methods [83]. Studies demonstrated that ML and DL models had been applied for EMG-based gesture recognition [84–86]. Tsinganos et al. employed convolutional layers and learning methods for hand gesture recognition and analyzed based on loss graphs and error analysis. The results demonstrated that a convolutional neural network (CNN) could be a successful method for EMG signals of hand gesture recognition, which classified 53 different hand gesture recognition [87].

Moreover, EMG biofeedback (EMG-BF) is a commonly employed technique to rehabilitate patients with muscular dysfunction, stroke, motor impairment, peripheral nerve injury, and hyperactivity disorder [88–91]. EMG-BF is a method, which displays a patient's electrical impulse on an external display device through audio–visual indications. EMG-BF is considered superior to classical EMG as it provides a recording of electrical signals in real-time [92]. In addition, the application of EMG in the rehabilitation of patients with cognitive impairment and locomotor dysfunction has been demonstrated from the past several years [93–98]. Liu et al. predicted neurodegenerative diseases (NDDs) in idiopathic rapid eye movement (REM) sleep behavior disorder. In the study, the authors concluded that EMG activity level acts as a biomarker of NDDs in REM sleep behavior disorder, where tonic EMG activity in REM sleep stably predicted the progression of Parkinson's disease (PD). However, phasic EMG activity level in REM sleep behavior disorder unstably predicted NDDs [99]. Similarly, Loconsole et al. successfully classified EMG-based and computer-vision handwriting analysis in the pathogenesis of PD [100]. With the advent of large EMG data, there is a growing need for critical analysis of this generated data to rehabilitate patients with neuromuscular diseases. Recently, Skelly et al. developed an open-source toolbox for analyzing and recording of motor-evoked potentials (MEP) extracted from transcranial magnetic stimulation (TMS) to evaluate the functional outcome of the corticospinal tract (CST) [101]. Similarly, Jackson and Greenhouse 2019 developed an open-source tool known as VETA: Visualize EMG TMS Analyze based on MATLAB to collect and analyze EMG data combined with TMS. VETA allows the presentation of behavioral stimuli from a single computer system [102]. Together, mounting evidence suggests that EMS is a powerful technique for collecting, analyzing, and visualizing neuromuscular electrical signals to rehabilitate patients. Several AI, ML, and DL models and tools have been developed for analyzing EMG data.

2.5. Electroencephalography

EEG is a non-invasive technique for measuring the brain's electric field through electrodes placed on the scalp, which records voltage potential due to current flowing in the neurons [103]. EEG is a powerful neuroimaging method with extended application in translational as well as in computational neuroscience. However, EEG does not measure the actual action potential of the neurons; instead, it measures post-synaptic potential [104–106]. Mounting evidence suggests the application of EEG in combination with ML algorithms in the detection of different neurological disorders such as epilepsy, Alzheimer's disease (AD), PD, schizophrenia, autism spectrum disorder, postanoxic coma, and posttraumatic stress disorder [107–116]. As illustrated in (Fig. 1(e)), where, the epileptic patient undergoes data preprocessing and feature selection followed by PBI analysis provided with high power. Afterward, the ictal period is calculated, and the information passes for a calculating distance of networks followed by identification of the starting point of seizure. Then, with the help of a correlational network, the therapist identifies the region of origin of the seizure. Emami et al. detected the seizure possibility using a CNN based on scalp EEG plot images. The results concluded that the performance of CNN (74 %) was higher than that of commercially available software such as BESA (20 %) and Persyst (31 %) [117]. Similarly, Naim-Feil et al. experimented with 21 schizophrenia patients and 28 healthy controls and demonstrated altered brain network dynamics in schizophrenia. The authors using Pearson cross-correlation algorithm and constructed static and dynamic networks in response to cognitive stimuli. The results demonstrated that patients with schizophrenia have dynamic network patterns in both global and regional areas [118].

Further, EEG has been successfully applied to detect and diagnose several neurological disorders associated with ML algorithms, where AD and PD were two prominent diseases [119,120]. Moreover, apart from detection and diagnosis purposes, ML approaches serve as a potential tool for the classification of EEG signals [121–124]. In addition, EEG is a powerful technique for the detection of characteristic waveform shape in PD pathogenesis. Jackson et al. observed the correlation between β oscillations of canonical shape and sensorimotor cortex, which can be helpful for the detection of PD progression [125]. Not only this, literature analysis observed the implication of ML in the classification and analysis of EEG signals in the treatment of several brain pathologies [126–131]. For example, De Taillez et al. experimented on 20 normal-hearing listeners to determine the listener's attention using non-linear neural networks, which achieved an outcome performance seven times higher than the linear baseline [132]. Moreover, the major drawback is the inaccessibility of electrophysiological signals from the cerebellum to MEG and ECoG. Thus, several attempts were made to increase the sensitivity of signals and found out that the congregation of EEG signals with ECoG and MEG for detecting and mapping the brain activity detected the signals with a suitable experimental setup [133–137]. Thus, with the growing scientific application of EEG for the detection, classification, and treatment of neurological defects, several algorithms, tools, and software have been developed to increase the effectiveness of the EEG method. Batch EEG automated processing platform (BEAPP) [138] and the brain imaging data structure (BIDS) [139] are two software for EEG, where BEAPP is an automated processing platform for EEG data, whereas, BIDS is an extension to EEG, which is used to analyze and share EEG data. Similarly, EngageMeter [140] for implicit audience engagement sensing using EEG data and EmotionActuator [141] for embodied emotional feedback through EEG were developed to increase the applicability of EEG for diagnosis of neurological defects. In addition, in 2017, BrainAtWork [142] was developed for analyzing the cognitive state of the person while working on a different task, which is based on EEG data, whereas, in 2018, Harvard automated processing pipeline for EEG (HAPPE) [143] was constructed for standardization of developmental and high-artifact data. Recently, in 2019, a novel software for measurement of EEG signal was developed by Teng Lin et al. [144], which is known as augmented wire-embedded silicon-based dry-contact sensor (WSBDS).

2.6. Wireless sensing networks

Recent trends demonstrated the extensive use of AR and VR technology in developing healthcare robots with WSN-based Internet of Things (IoT) [145]. WSN is a technology used to monitor patient's illness over a specific area with the help of a large number of sensor nodes. Each sensor node focuses on sensing some physical attributes such as temperature, humidity, pressure, and so on, which should be transmitted through wireless communication. In addition, each transmitting sensor node is capable of sensing either one or multiple attributes in a single-hop or multiple-hop manner and transmitted it to a single receiving node in the network called a sink node or base station (BS) [146]. The structure of a WSN is divided into three categories: star network, mesh network, and hybrid star-mesh network. In a star network, a single BS can send/receive a message to several remote nodes, where remote nodes cannot send messages to each other, whereas, in the mesh network, one node can transmit data to other nodes within its frequency range. In addition, in the hybrid-mesh network, the nodes with the least power consumption cannot send information or data to other nodes, which allows minimal power consumption to be maintained [147]. Recent advances in technology and especially in wearable sensing devices have emerged as a powerful resource with various applications in the detection, classification, and analysis of life-threatening cognitive disorders such as AD, PD, and associated gait [148,149]. The primary purpose of WSN-based devices is to improve the quality of patient's life, continuous monitoring of patients with dementia and cognitive disease, and assessment of sleep quality, which offer great avenues for advanced diagnosis and treatment. For instance, Bringas et al. constructed an ML-based CNN method to classify movement patterns on 35 AD patients. The results demonstrated that the proposed algorithm has 91 % accuracy in detecting AD stages [150].

Similarly, Moon et al. developed different ML-based models for classification of PD based on balance and gait characteristics from wearable sensor motions. The results concluded that the accuracy of developed ML models was more significant than the dummy model (0.48) or logistic regression (0.53) [151]. Not only this, several studies were performed for classification and analysis of degenerative disease using ML algorithms based on data from wearable sensing devices [152–155]. Monitoring patients with degenerative disease becomes crucial as patients' lives become miserable once after the progression of the disease. Thus, a different model system was developed based on WSN such as IoT sensor-based hand trembling robotic system, CGA methods and tool, SmartWalk BAN, Mobil4Park, flexible hybrid silicone-textile sensors, fog-based approach, Wi-Mind, cognitive wireless sensor networks, and medical cyber-physical systems (MCPS) [156–164]. For instance, Gugliandolo et al. developed a WSN-based system to record movements-tremor of patients with NDDs. The projected device is based on the array of four accelerometers, which are connected to an implanted development board. Further, the proposed device can monitor tremors/movements and accidentally falls after collecting the data through the internet connection or an external Bluetooth device [165].

Moreover, WSN-based wearable sensors with internet connectivity or Bluetooth connectivity have been successfully applied to quantify PD and essential tremor characteristics [166,167]. Thus, the evidence suggests that WSN emerged as a powerful technique for the classification, detection, and analysis of degenerative diseases such as AD and PD. However, WSN comes with limitations such as design issues and security issues. The design issues of WSN include fault tolerance, scalability, production costs, hardware constraints, sensor network topology, transmission media, and power consumption, whereas, security issues include confidentiality, integrity, authentication, and availability [168–170]. Thus, there is a growing need to eliminate the above-said constraints for better monitoring patients with life-threatening degenerative disease to improve the quality of life.

3. AR/VR technology as a potential tool for diagnosis

Many researchers have started exploring AR and VR technologies to detect neurological and brain disorders that are not possible with traditional diagnostic tools and techniques. For instance, Patients with multiple sclerosis (MS) have a great probability of falling and getting injured while walking, even when they think they can walk properly. It is very difficult to diagnose such dysfunctions, and they are usually detected very late, but recently a research team has shown that this problem can be detected at an early stage using VR. In their research, Selgrade et al. used VR to create an illusion that tricked the subjects under investigation to think that they were falling while walking on a treadmill [171]. In this study, the authors recruited 14 patients with MS and 14 healthy controls, and they were allowed to walk a treadmill while observing a VR environment using a VR device. During normal walking conditions, no major difference was observed in the walking style of MS patients and healthy controls. In contrast, once researchers created an illusion of falling in a VR environment, a significant difference was observed in step size, step width, stability of MS patients compared to healthy controls. Based on these results, one can conclude that this VR-based test can detect walking impairment in MS patients. Further, a healthcare company Altoida (https://www.altoida.com/), uses an AR-based application to diagnose cognitive dysfunction. The Altoida iADL test is an AR-based smartphone application that gives the user a variety of tasks in an AR environment, which tests the cognitive aptitude of the user through the XGboost algorithm [172]. This application observes how a user interacts with digital objects in an AR environment. The iADL test gives a series of everyday tasks, like hide and seek, in a gamified way, to the user, and the user interacts with digital objects in their physical environment. The application's AI algorithm observes the errors made by the user with the help of the smartphone's camera and accelerometer to predict the cognitive aptitude of the user. Further, Howett et al. have proposed a VR-based navigation task for early detection of AD. The authors have proposed a VR-based test in which a patient wears a VR headset and undergoes a navigation test while walking in a simulated VR environment [173]. In this test, a person whose entorhinal cortex region of the brain is fully functional will complete the test, whereas a person whose entorhinal cortex region is damaged will perform badly. The entorhinal cortex is a region of the brain that helps in navigation in real life, and it is one of the regions that is damaged at an early stage during AD pathogenesis. Hence, a person with AD performs poorly in this test, and thus this test can be used for AD diagnosis. In their study, Howett et al. took 45 patients with mild cognitive impairment (MCI) and 41 healthy control samples.

Patients with MCI performed poorly in the navigation test when compared to healthy controls. Out of 45 MCI patients, 12 patients had AD biomarkers, and their performance was worse than MCI patients without AD biomarkers. Furthermore, one study used, VR version of multiple errand test (VMET) developed using NeuroVR (http://www.neurovr2.org/) to analyze whether sleep dysfunction during early stages of PD affects decision-making abilities [174]. This study found out that patients with sleep dysfunction performed badly in the VMET test compared to controls, suggesting that sleep dysfunction affects decision-making abilities. [4]. In NeuroVR, raw data passes to an open vibe for data processing and classification, that further moves to VRPN connection for data preparation. The data is provided with an HTTP connection connected to the NeuroVR application, and finally, data is visualized through an external display device (Fig. 2(a)).

Additionally, Orlosky et al. have devised a VR framework for remote diagnosis of NDDs by evaluating eye movements linked with NDDs [175]. Here, Orlosky et al. have proposed a VR framework, which consists of a 3D VR environment simulated with the help of a VR headset with an infrared camera integrated into its lens to detect eye movements and to get high-resolution pictures of the pupil. In this framework, the researchers have created and designed tasks to stimulate eye responses in a VR environment. Tasks such as fixating on a point, smooth tracking of an object, and doing saccades used in the diagnosis of NDDs are emulated in a VR environment, then the camera in the VR headset analyses the eye movements takes images of eye response to different tasks. Then with the help of eye-tracking algorithms and image enhancement techniques, eye motion abnormalities can be detected. Hence, this whole framework can be used to detect eye abnormalities related to various NDDs. Moreover, NeuroDotVR is a neuro-ophthalmic diagnostic device that can be used to diagnose neurological disorders like MS, optic neuritis, among others [176]. NeuroDotVR consists of a NeuroDot brain sensor that measures visual evoked potentials and fields (VEPF) in response to a stimulus. In a way, NeuroDotVR is a portable, wireless VR display headset, which has a brain sensor to measures evoked potentials and scalp activity in response to a visual stimulus generated by the VR headset driven by a smartphone. As represented in (Fig. 2(b)), where, neuroelectric potential and EFEG rays from subjects were transferred to the NeuroDot device, which consists of an analog front end, vSync event system, USB port, and real-time clock. The USB port is connected to the NeuroDotVR server, which transfers information to WIFI. Inside the mobile application, NeuroDotVR visual stimuli generator renders the information for display.

4. Congregation of IoT, AR, VR, and neurological disease

IoT can be defined as an ecosystem of interlinked computing devices, hardware and software, digital instruments and people, with the facility to transfer pertinent data over a network and interact with one another [177]. In IoT, each of the objects has a well-defined and specific role to play, working in tandem. The main goal of IoT is to improve the efficiency of various services and systems and improve the quality of life. IoT in healthcare refers to the interconnection of various medical devices, systems that can collect, store,

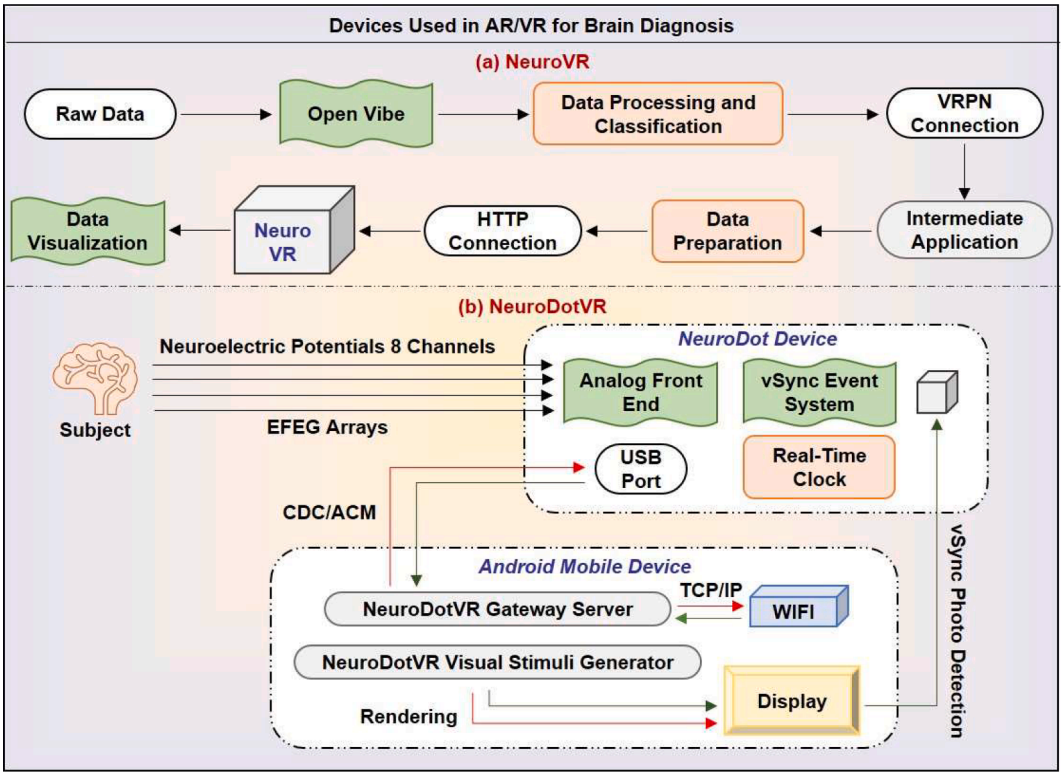


Fig. 2. Depiction of two distinct devices used in AR/VR technology for brain diagnosis: (a) NeuroVR and (b) NeuroDotVR.

analyze, and transfer many medical and biological data [178]. IoT is being used in the healthcare sector for different purposes like hospital management, telemedicine, patient monitoring, and home healthcare. Likewise, in neuroscience, various IoT-driven systems are being devised for patient monitoring, diagnosis, remote monitoring, assistance, and rehabilitation purposes. For instance, DeTrAs is a DL and IoT-based framework for diagnosing and assisting AD patients [179]. DeTrAs is composed of three main components: in the first component, input data is fed into a recurrent neural network (RNN) through sensors to verify that subject under investigation is an AD patient; in the second component, CNN and natural language processing (NLP) is used to determine abnormality stage of AD patients based on video and audio feeds; in the third component IoT devices are used to assist AD patients like helping in decision making (Fig. 3(a)). Likewise, another group has proposed an IoT-based application, referred to as iCare, to assist AD patients [180]. It has two main components: the first component requires the AD patient to wear IoT based Apple smartwatch; the second component is an iOS application that should be downloaded in the caregiver's or family member's smartphone and connected to the smartwatch of the patient to track patient's location, heartbeat, and other information. In addition, Sciarrone et al. have proposed IoT-based wearable smart glasses, that can analyze eye blinks, as many studies have reported a correlation between abnormal eye blinks and the onset of neurological defects [181].

Further, IoT, AR and VR, are the three most sought-after and swiftly expanding technologies. The congregation of IoT environments with AR and/or VR is a revolutionary idea. The use of VR and its immersive techniques in conjunction with the IoT environment has created a distinctive way of communication known as immersive transmission [182]. In a certain sense, this gives digital objects physical mannerism, and people can interact with them like they do in the physical world. Likewise, using AR techniques in conjugation with IoT environment offers a unique and semi-tangible interface for users to visualize and interact with IoT devices intuitively. Now, various systems and frameworks are being proposed which conjugates IoT with AR/VR. For instance, Park et al. have proposed VisIoT, a system which tracks the site and position of a wireless sender in IoT devices and displays the collected data in real-time, at the site of the sender, in AR device screens like smart glasses or tablets, thus empowering an intuitive interaction between the IoT device and the user [183]. In this VisIoT framework, a user explores and takes a video of its surroundings, and if it is found that the wireless transmitter is in the camera angle, then we ascertain the position of the transmitter in the camera image and superimpose the collected data over it.

Moreover, some researchers have proposed the congregation of IoT with AR/VR in the field of neuroscience. For instance, Ghorbani et al. have proposed integrating IoT with AR to aid patients with AD. In their study, the authors have proposed an AR-IoT-based system to aid AD patients, made up of two main components: the first component is an Android and/or Windows application that allows caregivers and family members of the patients to monitor the patient's condition at home and get a notification or warning if the patient is in danger; the second component of the system allows the patients to use AR headsets to scan quick response (QR) codes in their surroundings and get notification from IoT devices in the form of audio and 3D image [184]. In this system, authors have proposed an IoT-based android, windows application through which family members or caregivers can keep an eye on the patient status when the patient is alone and get a notification with the help of data collected by various IoT sensors placed at different locations

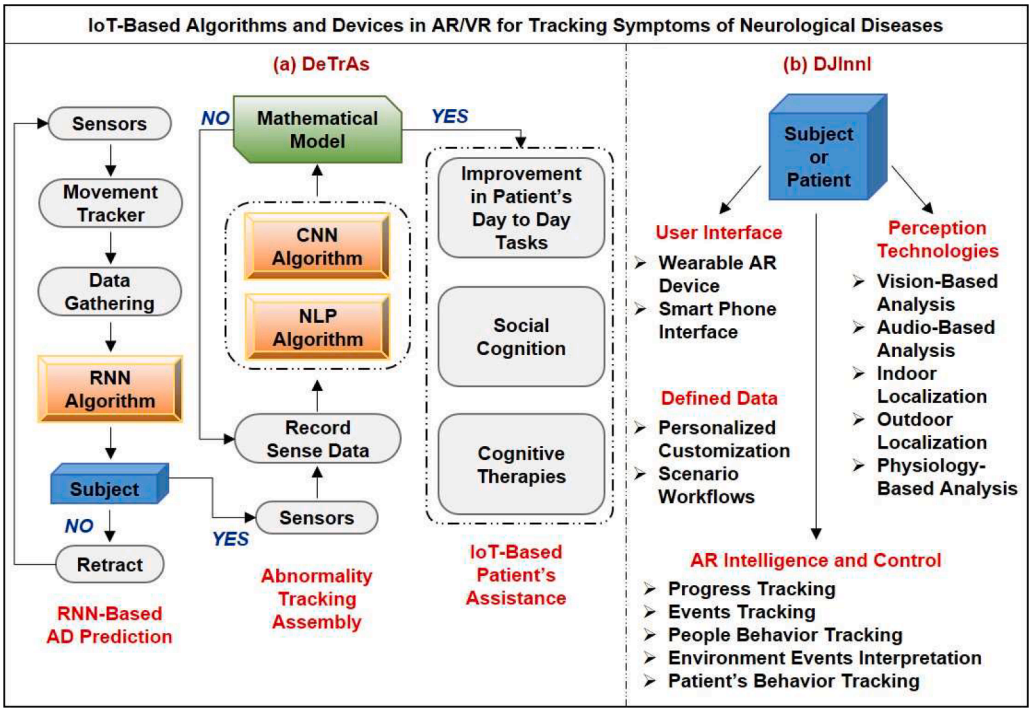


Fig. 3. Application of IoT-based algorithms and AR/VR technology devices to assist patients with neurological diseases: (a) DeTrAs and (b) DJInnI.

in the home. Through the smartphone application, family members can manage IoT actuators state in home, through this patient can get audio message or an image on AR screen from IoT device, reminding about different events like medication time and eating time.

Further, Rashid et al. have also proposed integrating AR and IoT to help people with motor disabilities. In their study, Rashid et al. have proposed an AR-IoT-based system that empowers wheelchair users with motor disabilities to interact with items placed far away from their arm's length [185]. This system is composed of radio-frequency identification (RFID), a touchscreen interface and an AR device. A wheelchair user can digitally interact with physical objects kept on a shelf beyond their arm's reach through this system, like in a library or a supermarket. Using this system, a wheelchair user can browse a shelf by clicking on it, explore all the objects it contains, selecting the desired object, and getting all the information related to it. The shelf used in this system is an RFID-based smart shelf which contains information of all the items kept on it. The user has to point the hand-held AR device towards the smart shelf, explore and select the desired item, and then overlay the information about the selected item in front of the user. In addition, Ben-Moussa et al. have devised DJInnI, an AR-VR-IoT-based therapeutic tool for people with social anxiety disorders (SAD) [186]. DJInnI conjugates AR, VR, and various IoT-based sensors to aid SAD patient's cognitive behavioral therapy, where the patients are analyzed through three different ways such as user interface and defined data, perception technology, and AR intelligence and control (Fig. 3(b)). DJInnI has two main components: AR exposure therapy (ARET) and VR exposure therapy (VRET). In ARET, DJInnI acts as an intelligent assistant for the user and counsels the user in a real-life environment. With the help of sensors, the system here explores the real-life surroundings of the user and based on pre-defined instructions given by the therapist, and the system displays information on the AR screen of the user to help him deal with the situation. The VRET component augments ARET by allowing the patient to recall

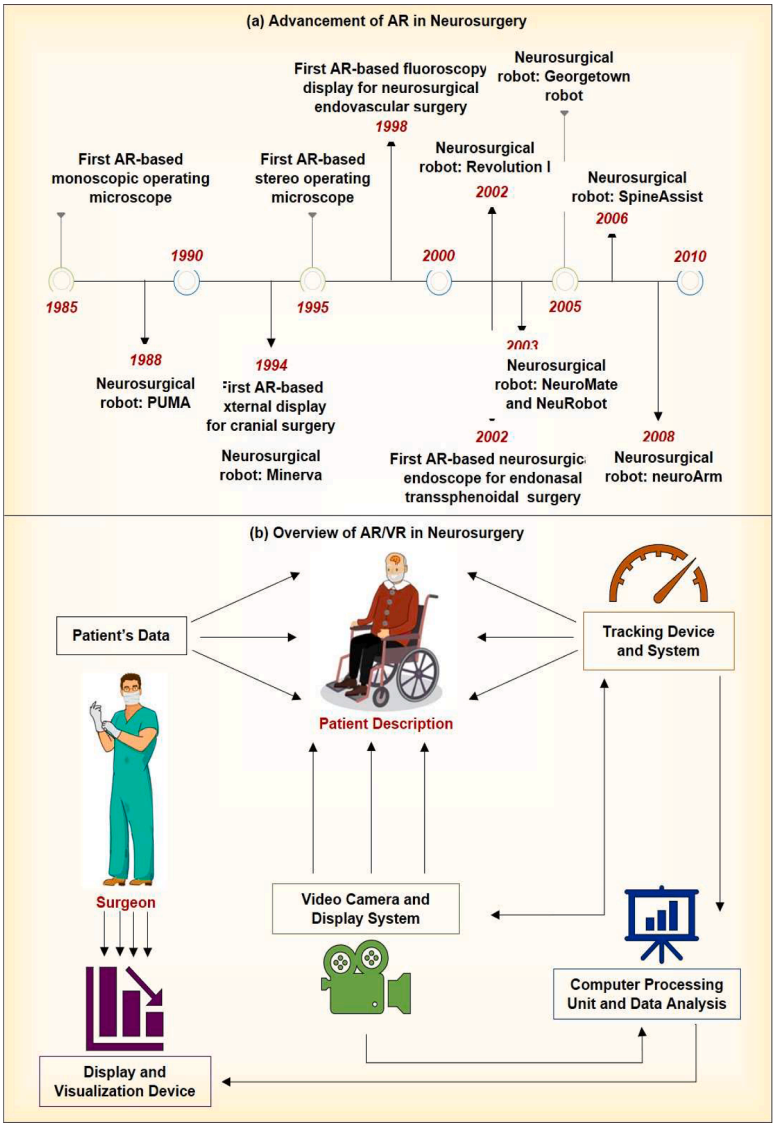


Fig. 4. Pictorial representation of advancement and overview of AR/VR in neurosurgery.

real-life situations in the immersive 3D VR environment. The VRET component analyzes the data collected by the ARET component in a real-life scenario and replicates a similar scenario for the patient in VR mode. Although the congregation of AR/VR with IoT has brought a revolutionary change, a massive amount of effort is still needed to properly exploit the potential of these three technologies.

5. Intricate role of AR/VR in neurosurgery

Neurosurgery is a field that has always been pushed to implement new technologies to advance and fasten the diagnosis, treatment and to provide long-term benefit to patients. It has been noticed that every year nearly five million people are unable to afford neurosurgery, particularly people living in low- and middle-income countries (LMICs) are majorly affected. Moreover, due to the COVID-19 pandemic, even established successful treatment options were not accessible. Thus, there is a desperate need to create virtual solutions. Immersive technologies, of which AR and VR are two principal types, are one of the budding applications and a promising approach for accomplishing global neurosurgical justness. AR can provide better visualization and teleproctoring, whereas, VR can help surgeons navigate the surgery and offer remote neurosurgical training [187]. In 1985, the first AR-driven microscope for cranial surgery was developed at Dartmouth [188]. Soon, in 1994, the first video augmented device was developed for external display [189]. The very next year, the first augmented stereo operating microscope was developed in the United Kingdom [190]. Augmented fluoroscopy display for neurosurgical endovascular surgery was developed in 1998 [191]. In 2002, an augmented neurosurgical endoscope was developed for endonasal transsphenoidal approaches [192,193]. Similarly, the neurosurgical robots known as Georgetown, Spine Assist, and neuroArm were discovered in 2002, 2006, and 2008, respectively (Fig. 4(a)).

Proprio, TrueVision, and Surgical Theater are three companies using VR and AR technology to improve neurosurgery. Advanced medical technology is one of the major issues behind constantly increasing costs in the healthcare sector. However, the current practices involving AR in neurosurgery are difficult to assess as many practical applications use AR system prototypes [194]. An article stated that with the help of 'Simulation Alliance,' where AR/VR technology and experience gained by using this technology can be shared among institutions, thereby reducing the financial burden and thus promoting the benefits and utility of it in neurosurgery [195]. The overview of AR/VR in Neurosurgery (Fig. 4(b)) states how the patient's data are collected from the patient, which is connected to a tracking device, display system, and computer processing unit. The data from the display system and tracking devices pass to the computer processing unit, which further transfers data to an external display device through which the surgeon monitors the patient. The following section will describe the application of AR and VR in different types of neurosurgeries.

5.1. Neuro-oncology surgery

Neuro-Oncology is an important and most frequent type of neurosurgery that pacts with the management and treatment of brain tumors. Efforts are being made to limit its invasiveness. The major issue is the infiltration of malignant brain tumors into the skull, making the resectability of tumors from normal brain tissue extremely difficult. Moreover, because of the small field of vision, slight damage to normal tissue can lead to an irreversible impairment [196,197]. Presently, neurosurgeons rely more on image-guided neurosurgery (IGNS) to resect the brain tumor based on preoperative images of the patient's brain to track the tumor [198]. However, it has been posited that the plan of neurosurgeons gets disturbed; firstly, due to continuous switching sights between the computer screen and the surgical field. Secondly, as the computer screen gives three separate images representing different anatomical planes, the surgeon has to merge all these to make a 3D image in their mind, which hinders their workflow. A solution to this problem is a specific modality named NeuroPlanner, a stereotactic brain atlas that let neurosurgeons map, plan and modify the procedure in real-time within a mixed 2D–3D simulation [199,200]. Some prototype systems that are integrated with AR helps neurosurgeons in preoperative planning and take the best approach to remove the tumor completely from the brain. For instance, Fabrizio et al. developed an AR neuronavigator that is intuitive, easy to use, and can be effective in guiding skin incision, craniotomy and lesion targeting, also, their testing platform can be used to provide training in brain tumor resection procedures [201]. Deng et al. used a wireless tablet computer AR neuronavigation system for surgical planning in China [194,198]. Inoue et al. made an AR system prototype where a web camera was used to superimpose the image of a brain tumor over the patient's skull and dura [202]. Mahvash et al. developed an AR system that integrates magnetic resonance (MR) and CT of a patient's brain that would further help neurosurgeons to improve their approach in surgery [203]. Likewise, the intraoperative brain imaging system (IBIS) is an open-source simulation platform that uses AR to provide the inaccuracies in surgery by identifying the irregularities between preoperative imaging and intraoperative ultrasound [204]. Moreover, a collaboration between Brainlab and Magic Leap has given rise to an AR technology that can be used intraoperatively in skull-based surgery [200]. It has been found that AR-based surgery helped in getting volumetric CT and MRI data in no time and reduced the hospital length stay by 40–50 % in comparison to cases that were not operated on using AR technology; this was proved in one series, which consisted of 74 patients in which 64 had primary or recurrent gliomas [205].

5.2. Neurovascular surgery

Neurovascular surgery is another subfield of neurosurgery which deals with the treatment of patients with cerebrovascular disease. As surgeons rely more on images to make intraoperative decisions, it is said that neurovascular surgery is strongly dependent on the augmentation of stereomicroscopes. The main advantage of AR stereomicroscope is that it takes only additional 20 min to create a virtual environment. Images captured by microscope include either fluorescence images from intraoperative CT angiography (CTA), magnetic resonance angiography (MRA) or digital subtraction angiography (DSA). Firstly, to create a virtual environment, registration of every component of the patient is done via a standard IR-LED-based neuronavigation system.

Secondly, verification of the registration is done at each step taking from skin incision to arachnoid dissection. Moreover, in extracranial–intracranial bypass procedures, AR technology has proved superior to manual pulse palpitation and is quite equivalent to Doppler ultrasound or intraoperative DSA-guided donor vessel identification [206]. Arteriovenous malformation (AVM), a congenital disease, is one example of a cerebrovascular disease where the hallmark feature is the intertangling of vessels in the brain or spinal cord. As these vessels are thin-walled, they might rupture and bleed, resulting in a stroke, thereby causing death [207]. Microsurgical resection is an established procedure currently practiced, which involves performing craniotomy and resecting the AVM. AVM surgery uses a magnifying microscope to enlarge the area where the surgery is to be perform. However, it does not deliver the necessary information required to aid the surgeons to identify these blood vessels, thus making the whole procedure complex, resulting in longer operative time. A study stated development of an AR system prototype that could offer MR visualization during AVM surgery as the system incorporated chromadepth feature, which provides the distance of vessel from brain surface and it also provides color-coding, which helps surgeons to better visualize arterial and venous blood vessels under the microscope [194,208]. The 3D imaging tool is also being used in invasive angiographic procedures to treat brain aneurysms as it requires surgical management because of its high mortality and morbidity rate [209]. In 2016 with the help of Immersive Touch technology, a real-time sensory haptic feedback virtual reality aneurysm clipping stimulator was developed, which was tested by neurosurgeons, and it came out extremely beneficial as it not only provides simulation indicating real-life surgery but would also provide preoperative surgical rehearsal with realistic haptic feedback [200,210]. Neurosurgeons have also used virtual interactive presence and AR (VIPAR) to guide remote carotid endarterectomy [211]. Moreover, AR technology has correspondingly helped in getting volumetric intracranial CTA data which could be overlaid onto a real-world video feed [212].

5.3. Spinal surgery

As many surgical procedures are moving towards a less invasive approach, advancement in AR/VR applications has attracted a lot of attention towards spinal surgery compared to other surgical specialties. Although sparse data is available for AR applications in open spinal procedures, AR technology still holds a promise to provide real-time trajectory guidance for percutaneous instrumentation or at superficial to skin level [213]. Even though the present stereotactic navigation systems help guide the hardware trajectory compared to bony anatomy, leave the skin projection of these trajectories at the surgeon's decision [193]. For the surgery to be successful, simulation training is required to minimize the risk to patients. The current VR modalities for simulation are pedicle screw insertion, vertebroplasty, posterior cervical laminectomy and foraminotomy [214]. A study stated that pedicle perforation for navigated screw insertion has a significantly lower risk than non-navigated insertion for all spinal regions [215]. Currently, C-arm intraoperative fluoroscopy is typically used to direct percutaneous instrumentation. A technique has been stated that utilizes AR technology which involves placing a standard video camera that virtually coincides its optical center with a C-arm x-ray source, which further helps to overlay x-ray and real-world images [216]. Interestingly, it has also been stated that in comparison to the C-arm technique, an AR-based modality, holographic 3D AR-assisted guidance has shown a reduction in needle passes to 54.2 % and radiation exposure to 41.2 %. Moreover, it also decreases the procedure time for CT-guided imaging [217].

Moreover, a prototypical AR application developed using MR glasses, Microsoft HoloLens, suggested high accuracy and a reduction in orientation errors [218]. Further, a cadaveric study projecting 3D reconstructed spine CT imaging over the cadaver torso stated that AR projection alone gives the maximum target error of 9.1 mm as it cannot target the final needle positioning because of the unavailability of angular information. However, AR combined with C-arm works better with an entry point error of 4.4 mm [219]. Thus, multiplanar cross-sectional imaging can provide the angular information required to target the implants to the desired transpedicular position accurately. There are very few industries leading in spinal surgery like Immersive Touch Inc, Novint Technologies, Medtronic plc, and Osso VR [220,221].

5.4. Training in neurosurgery

Surgery nowadays is moving from the Halstedian apprenticeship model, i.e., training on live patients to complementary training using simulation environments. External ventricular drain (EVD) is a common surgical procedure that junior residents in an emergency setting do, and as to do it, they have to refer to axial CT-scan of the patient's head; and undesirable trajectories can lead to multiple attempts, which potentially land up to 14–20 % high chances of getting misplaced drains, therefore, increasing the risk of hemorrhage and infection [222]. AR and VR are the key instruments in neurosurgical training and can create an advanced learning environment compared to traditional pedagogic schemas. It provides the resident physicians and neurosurgeons with the opportunities to master surgical techniques and provides spatial retention, resulting in greater efficiency, fewer technical faults, and thereby, minimize the risk factors that the patient undergoes and thus benefit a diverse patient population. Surgical planning is an area of interest that uses VR to aid in preoperative planning, which also pushes towards a less invasive approach and microsurgical procedures that allow the surgeons to generate a 3D reconstruction of 2D tomographic imaging, which includes CT, CTA, MRI, and MRA. This further helps surgeons to manipulate the anatomical structures in stereoscopic format with the help of 3D tools in a virtual display [223]. VR systems loaded with patient-specific images can help neurosurgeons practice a forthcoming procedure; this will act as pre-planning that can give surgeons a sense of familiarity and save time during the actual operation [224]. With the addition of software tools, surgeons can modify the threshold, color and transparency of the images. This approach has been shown to reduce mortality rates to 0 % and morbidity rates to the single-digit percentile for intracranial aneurysm and subarachnoid hemorrhages [225,226]. The 3D learning environment can help improve the knowledge representation, contextualization and technical abilities of the learner [227]. In neurosurgery, VR modalities are immensely treasured for education because of the complicated and complex neurosurgical

procedures. Strict and formal training can help budding neurosurgeons to advance their skills. Moreover, evaluation criteria can be set for both naïve and skilled surgeons [228]. VR simulation curriculum can also help set different levels of education and skill requirement in neurosurgery via objective assessment and detect individual strengths, weaknesses, and others areas of skill improvement for neurosurgeons. This will help in maintaining and tracking the skills acquired by residents and senior neurosurgeons [229]. Currently, AR modalities are in primitive stages, although exciting potential has been shown towards surgical practices like laparoscopic visualization of internal organs [230]. Moreover, it has been shown to increase precision in laparoscopic procedures [231]. AR technology has also been used in cerebral aneurysm clipping, AVM, tumor resection and shunt placement [232]. Quite a few AR modalities have also shown specific uses in the field of endoscopic neurosurgery [233–236]. The dynamic AR display can be presented on head-up displays, HMDs, intraoperative microscopes and projections. Moreover, with the live-mapping AR, surgeons can rely less on HMDs and can integrate virtual processing directly onto the patients so that they can maintain their focus within the surgical field [237]. Lastly, AR technology help neurosurgeons to see hidden structures intraoperatively without the surgeons taking off their eyes from the surgical field [228,238,239].

6. Application of AR/VR in neuromodulation and neuroprosthetics

6.1. Neuronavigation

Due to its ability to superimpose 3D images onto the real environment, AR has been continuously explored as a potential alternative to conventional neuronavigation systems. The most common application of AR-based neuronavigation systems is neurosurgical procedures, stereotactic planning, and stereotactic guidance [240,241]. For instance, W. Deng et al. developed a wireless tablet computer system based on AR for observing the intracranial anatomical structure of the patient during surgical procedures. The designed system can transmit navigation information to a dynamic computer system through a local area network, which sends the information to a display monitor [198]. Similarly, Inoue et al. applied AR-based neuronavigation systems for clinical application to plan skin incision at three intervals before the operation, during craniotomy, and after dural incision. The authors included three brain tumor patients as subjects. The designed system was equipped with open-source software called a 3D slicer, the infrared-optical tracking sensor, and web cameras [202]. Recent years witnessed AR-based endoscopic skull application in modern surgery, especially in tumors for neuronavigation [242,243]. For instance, Gerard et al. combined interoperative ultrasound imaging methods and AR visualization, which improves image-guided neurosurgery [244].

Moreover, different studies demonstrated the combination of smart glasses and HMD with AR system for neuronavigation [245–247]. For instance, Cutolo et al. demonstrated that AR systems eliminate the shortcoming of commercial neuronavigation systems. The authors applied a two-phase evaluation process in a stimulated small tumor resection adjacent to Broca's area, where phase I involved nine patients without neurosurgical training and in phase II three surgeons were involved in evaluating the efficiency of the AR-neuronavigator. The results concluded that phase I revealed the ability of the AR scene to evoke depth perception under

Table 2
AR/VR-based list of systems and devices used to assist cognitive dysfunction, gait impairment, and neuronavigation.

Device or tool	Technology	Description	Reference
(A) Cognitive Impairment and Gait Abnormalities			
NAVIG	AR	It is based on bio-inspired vision system used for recognition and locating the objects in a short duration. It also consists of bio-rendering system, which is used for identifying the location of the object.	[253]
ARCoach 2.0	AR	Optimizing a vocational prompting system based on AR for people with cognitive impairments. The proposed system uses cyclic codes to enhance AR-tag recognition rates.	[254]
CogAR	AR	The App can be a useful tool able to provide more independence to patients during actual daily living activities and in navigating their home, providing cues for memory integration in real-time	[255]
ECO-VR	VR	ECO-VR evaluates temporospatial orientation, memory, language and executive functioning	[256]
cARe	AR	cAReis designed in such a way that it could assist the older individual with memory dysfunction in performing daily tasks	[257]
Scratch 2.0	AR	Creates real world and virtual reality displays at the same time	[258]
Virtual Supermarket (VSM)	VR	It is designed to assist older adults with mild cognitive dysfunction	[259]
kayak program	VR	It a tool, which is created to improve memory and cognitive functions of elderly people	[260]
V-TIME	VR	It is a programme developed to minimize the risk of fall in elderly people	[261]
VMET	VR	It is a bioinformatics-based tool used for early detection of PD	[262]
Heel2Toe	AV	Gait training using the Heel2Toe sensor is feasible and potentially effective for improving gait quality in people with PD	[263]
(B) Neuronavigation			
Trans-visible navigator	AR	It is used in clinical investigations, where surgeons can easily view the surgical images with the help of augmented reality.	[264]
StealthStation™ S8	AR	The system is capable of receiving additional visual information of the structures being manipulated, increasing the precision of the surgical procedures.	[265]
HMD	AR	XVision is a recently FDA-approved head-mounted display for intraoperative neuronavigation system, which is wireless, having an integrated surgical tracking system and transparent retinal system.	[266]
MARIN	AR	It allows user to perform daily tasks through interaction between real-time data and augmentation data.	[267]
iPhone-assisted system	AR	Solution for the Localization of Intracranial Lesions	[268]

different visualization, whereas, phase II confirmed the potentialities of the AR-systems in assisting neurosurgical procedures [201]. In addition, AR-based neuronavigators were also applied for the treatment of neuropsychological diseases such as depression [248]. For instance, Sathyanarayana et al. demonstrated a comparison between different head pose tracking methods for transcranial magnetic stimulation, which is the most common treatment for depression. The authors in their study incorporated marker-based tracking with an MR headset (OST-HMD) camera, marker-based tracking with a world-anchored camera, and markerless RGB-depth (RGB-D) tracking with a world-anchored camera. The results concluded that all three methods attained the required accuracy for treatment. However, for smaller head rotations, only marker-based and marker-less methods obtained the required accuracy, whereas, for larger head rotations, the marker-based method achieved the required accuracy [249].

VR systems are also applied for neuronavigation during the surgical procedure, but the evidence is less than AR-based systems. Two VR-based systems for neuronavigation are SonoWand [250] and EasyGuide Neuro™ [251], which are used for ultrasound-based neuronavigation and the course of brain shift during microsurgical resection, respectively. Further, Stadie et al. performed a study, where the authors enrolled 48 patients, which underwent an invasive cranial surgical procedure to demonstrate the distinction between VR planning system and intraoperative image-guided system. The results demonstrated that the VR-based neuronavigation system correctly determined the craniotomy's size and position in 47 patients, whereas, in 44 patients, the proposed system correctly identified the craniotomy site [252]. Although different AR and VR-based neuronavigation systems have been applied successfully across many neurosurgical procedures, more research is required to extract the complete efficiency and accuracy in registration. Further, studies are required to assess whether AR or VR-based systems for surgery are safe and effective for the widespread population. The adaptive response of the patient towards AR and VR-based assistive systems should also be studied (Table 2).

6.2. Episodic memory

Episodic memory is a type of long-term memory associated with recollection of past experiences and past events. It is one of the memories which gets affected during neurological disorders. It has been observed that VR-based tools can be used for the assessment and remediation of episodic and other long-term memories. For instance, Bréchet et al. used immersive VR to prevent episodic memory loss [269]. In this study, subjects were immersed in a VR environment depicting two real-life outside scenes with the help of HMD. In one scene, subjects were in first-person perspective form, where they could see their hands, legs, whereas, in the other scene, the subjects could not see their body. After getting familiar with the VR environment, a ball emerged in the two outdoor scenes and the subjects were asked to track and follow the positions of the ball. Various daily-life objects were also kept in the two outdoor scenes to mimic a natural, real-life environment, and the ball moved in such a way that allowed subjects to notice each and every object kept in the two scenes. After one hour of this encoding session, the subjects were re-immersed in the two outdoor scenes, but this time some new objects replaced the old objects and later, the subjects were asked whether the two scenes had any changes or looked like they saw them previously. The results showed that subjects differentiated between two conditions with good accuracy when they were present in first person perspective form compared to no body condition. Further, Corriveau Lecavalier et al. devised a VR-based, Virtual shop for assessment and remediation of episodic memory [270]. 3DVIA Virtools 5, 3D engine was used to create a virtual environment for the virtual shop, and Nvisor ST50 audio–visual headgear was used for immersion in the virtual environment. The virtual environment mimicked a real-life convenience store where the subjects can move around and fetch required items. The subjects were shown a list containing 12 items used in real life, such as milk, butter, belt, and subjects were asked to memorize all the items. Later, subjects were asked to explore the store and fetch the items in the list, with the help of remote control to select and fetch the required items. The results showed that subjects had high motivation to explore the virtual shop, and their performance in the virtual shop was similar to traditional memory tests in real life. Likewise, Parsons et al. used VR-based, virtual environment grocery store (VEGS) to assess episodic memory in the elderly [271]. VEGS is a non-immersive VR environment that mimics a grocery store and contains a variety of cognitive assignments. The authors here placed participants inside the VEGS environment, where the participants had to memorize items on a shopping list showed to them, then explore and fetch all the items in the list while ignoring the fake items. The results showed that the scores obtained in the VEGS test were similar to traditional neuropsychology tests. Further, Wais et al. have devised Labyrinth-VR, a VR-based game to improve long-term memory [272]. In this game, the user dons a VR headset, then learns and navigates through virtual urban and rural neighborhoods to complete assigned tasks. The complexity and difficulty of tasks increase as the game progresses. While playing the game, players have to move their bodies around, thus getting physical exercise which increases cerebral blood flow. Further, Clemenson et al. [273] performed a study and reported that brain stimulation by a 3D virtual environment improves hippocampus activity and memory. In their study, two groups of naïve video gamers were trained on two different games for two weeks. Control group was trained on 2D angry birds' game whereas experimental group was trained on 3D super Mario game in a virtual environment. The results showed that video gamers in experimental group performed better in memory recognition tasks compared to control group.

6.3. Neuroplasticity and neurogenesis induction

Neuroplasticity can be defined as the capacity of the nervous system to reorganize itself and its functions after an injury [274]. It has been observed that VR-based therapy and training can promote neuroplasticity and neuronal reorganization. For instance, one study reported that VR therapy promotes cortical plasticity and reorganization in a patient with hemiparetic cerebral palsy [275]. Hemiparetic cerebral palsy is a neurological disorder that is associated with delayed motor development. In this study, 8 years old male with hemiparetic cerebral palsy had undergone VR-based therapy to check the effects of VR therapy on neuroplasticity and motor development. IREX VR system, which uses a video camera, LCD monitor, cyber gloves, was used for VR-based therapy. A video camera

is used to track user movements and immerse the user in the VR environment, where games like soccer and bird ball are played, promoting movement, mobility, and strength. The subject under investigation played each game five times for four weeks. Cortical reorganization of the subject was noted before and after the VR therapy, and it was observed that VR therapy induced significant neuroplastic changes at the sensorimotor cortex due to improved motor skills. Further, Wang et al. used leap motion-based VR system for neural reorganization and motor development in subacute stroke patients [276]. In this VR system, the patient can see the virtual image of his/her own upper-limbs while carrying out the routine tasks involved in VR games by controlling the virtual hand through his/her own limbs in real-time. The authors involved 26 individuals with subacute stroke in the experimental group, which received leap motion-based VR training as well as conventional rehabilitation for 4 weeks. In the VR mode, games like piano playing, object catching, petal picking was played. On the other hand, the control group received only conventional rehabilitation training. The results showed neural reorganization in fMRI and recovery of motor functions in the group which received VR-based training. Further, Xiao et al. used VR-enhanced treadmill training for patients with stroke to facilitate cerebral reorganization and motor development [277]. In this study, eight patients with ischemic stroke received VR enhanced treadmill therapy for three weeks. A virtual environment was created on a television monitor placed before a treadmill, which simulated various real-life scenarios like street crossing, lane walking, among others, while the patient walked on the treadmill. Post-training, the fMRI results showed cerebral reorganization in the patients, which was associated with improved walking and motor functions. Further, Anopas et al. have devised a VR-based gaming system to enhance motor skills and neuroplasticity [278]. The game was developed using a unity game engine, and scripts were written with the help of Javascript. In this game, a user has to hear a sound from an animal model and a thunderclap sound from a specific goal. The user has to identify and memorize the direction from which sounds are coming to damage the animal models by shooting and moving towards the goal to complete the game. The authors in their study compared the performance of their VR memory game with a common commercial game and found that subjects showed higher brain activity after playing their game compared to the commercial game.

6.4. Neuroanatomy

Recently, researchers have taken the help of AR and VR technology to study the anatomical structure of the brain. 3D visualization of the brain can help gain new insights and help with educational purposes, like learning and training. In this regard, NitLabEduca is an AR-based mobile application for visualizing and studying the spinal cord, its anatomy and physiology, and related pathways in an interactive way [279]. NitLabEduca allows visualization and analysis of the spinal cord, its related pathways, and physiology through an interactive 3D model of the spinal cord in an AR environment. NitLabEduca uses 3dS Max software to design the spinal cord in 3D, and during its implementation, the vuforia engine was used to create a mobile AR environment with codes written using C# language. NitLabEduca has a print 3D feature that provides a 3D model of the spinal cord in an STL format file, which can be printed using 3D printers. Further, The Holobrain (https://hive.med.ubc.ca/projects/holobrain/) is an AR-based tool for visualization of the brain and its anatomical structures in 3D help of a device like hololens. Holobrain creates 3D models of the brain by volumetric reconstruction from 2D MRI images. One of its main applications is teaching brain anatomy to students in an exhilarating way. Further, Souza et al. have devised a VR-based game for enhancing teaching and learning in the field of neuroanatomy [280]. The authors have proposed a puzzle-based game that tests cognitive aptitude, logic, strategy, and pattern recognition of players. Their game consists of a virtual environment where different portions of the encephalon like cerebellum, frontal lobe, temporal lobe, occipital lobe, corpus callosum and others, are presented in 3D forms in front of the students along with their names and functions. The challenge for the students is to assemble and match these parts in their specific locations in the shortest time possible. The user can take the help of the Oculus touch controller to mimic a virtual hand which can pick different anatomic parts of the encephalon and places it in the desired location. The game was developed using a unity3D engine and used an HMD for experiencing the VR environment. Further, Kim et al. used a unity3D engine and oculus rift to visualize a 3D model of the human brain [281]. The authors here loaded DTI tractography images of the brain in the unity3D engine through different methods then used the oculus rift to visualize the model obtained from the unity3D engine. Moreover, Kockro et al., conducted a study with a group of students, which showed that stereoscopic digital 3D models of the brain presented in a VR environment could augment neuroanatomy learning, as the students rated the 3D method of teaching better than the traditional 2D method [282]. In this study, the DextroBeam system was used, high-resolution 3D neuroanatomical models.

7. Bottlenecks in the applicability of AR/VR technology

The application of AR and VR-based technology in the field of neuroscience has come up with different shortcomings during its expansion. VR has been used to simulate external reality by using predictive coding mechanisms that allow people to feel ‘real’ but what is not there is the environment. VR tools can propose new ways for structuring, augmenting, and replacing the experience of the body for clinical goals. Besides, they can offer new ways of assessing the functioning of the brain by directly aiming the methods behind real-world behavior [283–286]. However, VR can stimulate the external body and not the internal body. For instance, it can help reproduce the features of the body through vision and hearing but cannot reproduce other senses like touch and smell. It can partially offer proprioceptive features of the body using haptic technologies but cannot offer interceptive features of the body [287–289]. Moreover, as the advancements in the field of behavioral and physiological neuroscience growing exponentially, our understanding of neural connection with social cognition increases, and thus the use of VR for research in social neuroscience is raising ethical concerns as experiments conducted using VR with a foreseeable consequence can cause serious and lasting harm to the subject [285].

As AR is in its primitive stage, it still has a role to play in the neuroscience area, particularly in psychotherapy and neuro-rehabilitation. Presently, HMDs are the most common devices to display AR and VR worlds. However, a few limitations that need

to be resolved are their weight, ergonomic features, and some technical parameters like resolution, the FOV, registration, and occlusion. Contrary to this, it has been observed that OST displays have an advantage over HMDs as they rely mainly on optical gear that includes the human eye. Besides that, the features of virtual studio technology are even more appealing than OST.

The shortcomings of AR which have been displayed by the above-mentioned devices were solved by technologies like projection display and VST, although, these two also have some limitations [290]. Besides the display, there are issues with positioning. Even after posing some obvious limitations, HMDs are more widely used, nevertheless, handheld devices are also becoming increasingly popular and are getting a lot of attraction. There is a promising prospect that the upcoming technology will include a combination of handheld devices with projection displays to ensure that the hands are available while AR is in use [290,291]. The use of AR in neurosurgery has also posited a few barriers as problems with system latency, image alignment, registration accuracy, and issues with tissue/organ movement [194]. Also, profound implementation and maintenance costs are a barrier to the adoption of VR and AR in neurosurgical practice [200]. Further, time codes might be difficult while synchronizing an AR/VR system with the true impact apparatus as they are not electronically linked. As there is a limitation in the diversity of test subjects, therefore, preparing a list comprising of complete physical constraints like age, gender, and hair type might be difficult. Also, a major hindrance is that the AR and VR environments might not be as immersive as the real world, and thus, the reaction or response time from the test subjects might not be as correct as real-world tests [287]. Besides these limitations mentioned above, the increased cost of advanced VR and AR systems remains relatively high.

Nevertheless, with advancements in computer equipment and an increase in computational power, there will be a gradual decrease in the price. Moreover, creating VR and AR worlds requires specialized skills in 3D modeling, texturing, and programming. However, because of the availability of powerful tools -some at no cost- has simplified these tasks. Lastly, cybersickness is a lingering concern as users who have repeatedly used VR or AR for therapeutic or training applications have reported experiencing nausea [290,292]. Still, these technologies continue to grow. However, the current state of literature states that more evidence is required to call this technology worth the time, cost, and effort.

8. Conclusion and future perspectives

AR and VR have finally come of age for grave applications in the field of neurosciences. The first article that discussed the implementation of VR in behavioral healthcare was published in 1995. More than 20 years later, the clinical potential of both VR and AR technology has come out as a huge advantage in diagnosing and treating various mental disorders [1]. Many authors of different studies have suggested that alteration in the function of the body matrix or other related processes could be the cause of different neurological or psychiatric disorders. However, one cannot deny that the incoming of VR and AR technology has opened a new trans-disciplinary research field named embodied medicine. According to neuroscience, our brain generates an embodied simulation that helps our body to predict necessary actions and thus regulate and control our body in the world. The VR and AR work in the same way as they try to predict the sensory information of the person's movement and provide him/her a sense that he/she is in the real world. Therefore, this technology can be well-defined as embodied technology, for it can modify the embodiment experience of its users. This technology has impacted many stakeholders. For instance, it has drawn the attention of many researchers who are working in cognitive and behavioral neurosciences, ethics and law, and mental health. Also, many rehabilitation and mental health practitioners are interested in using this technology in therapies.

Further, it has attracted many developers to use VR, AR and MR in their neuroimaging and monitoring tools. According to the "2020 AUGMENTED AND VIRTUAL REALITY SURVEY REPORT" by Perkins Coie LLP, the global healthcare industry is expected to grow nearly $11 billion by 2025. Nonetheless, it has also been stated that besides the entertainment and gaming sector, 38 % of the healthcare and medical devices sector will be most disrupted by immersive technologies in the next 12 months. Interestingly training simulation, which includes training of surgeons in the healthcare sector itself, is expected to flourish 68 % in the coming 2 years [293]. It has been found that VR has more applications in simulation, diagnostics, exposure therapy, rehabilitation, and pain distraction, whereas, AR applications are more focused on surgeries, rehabilitation, training, and medical education. Under VR, the diagnostic segment, and under AR, the training and medical education segment are anticipated as the fastest growing segments with a lucrative growth rate [294].

Nevertheless, at present, some critical shortcomings limit its possibility in the field of neuroscience. Firstly, VR and AR simulate the body externally and not internally. However, recently, a study introduced the concept of 'sonoception,' which is a non-invasive technology where wearable acoustic and vibrotactile transducers can stimulate the body internally, including the stomach, heart, muscles, and otolith organs of the vestibular system [295]. VR and AR have shown great significance in clinics, tools for assessing cognition, different therapies and rehabilitation for various mental illnesses, and training physicians and surgeons. However, the major question is whether VR and AR-based tools are better than currently present medical practices. Moreover, suppose it is true that these VR and AR applications help clinicians, therapists, and other caregivers. In that case, it is also true that some of these technologies will never have a large audience as their technical knowledge is just too ambiguous for personal use, although, with proper training, this limitation can also be well-tackled, besides that, as the production increases, price of the technology will decrease over time. Currently, developers of VR and AR are using handheld devices and eye-tracking devices to take intuition-based inputs. Also, further advancements in tools for mind-machine interfaces such as EEG and EMG as well as growth in AI and ML, are opening new avenues for mindreading or simply controlling input by thoughts. The main challenge that needs to be tackled here is to achieve sufficient resolution and discrimination between conscious or intentional actions versus subconscious or unintentional actions [296]. VR and AR are flourishing fields, but their application in neuroscience sets some priorities that should be the next step for this technology to go in the right direction. Firstly, better bridging is required amongst VR/AR research communities to build a better understanding of the

context of the feasibility of this technology amongst researchers, developers, and content creators. Secondly, VR, AR, and MR generate a lot of neurological data; there should be some way to access this data while protecting user privacy, data ownership, and research ethics. Lastly, it is very unpredictable to judge the users' social behavior on coming into contact with VR or AR applications. In summary, we encourage tackling the existing limitations and promoting future developments of VR and AR in the field of neuroscience, henceforth, intending to motivate and facilitate all the professionals working in this field to continue to contribute to this topic.

CRediT authorship contribution statement

Mehar Sahu: Data curation, Formal analysis, Writing – original draft. **Rohan Gupta:** Data curation, Formal analysis, Writing – original draft. **Rashmi K. Ambasta:** Project administration, Supervision, Writing – review & editing. **Pravir Kumar:** Conceptualization, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

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

Acknowledgment

We would like to thanks the senior management of Delhi Technological University for their constant support and guidance.

References

[1] D. Reiners, M.R. Davahli, W. Karwowski, C. Cruz-Neira, The combination of artificial intelligence and extended reality: a systematic review, Front. Virtual Real. 2 (2021) 1–13, https://doi.org/10.3389/frvir.2021.721933.
[2] L. Cen, D. Ruta, L.M.M.S. Al Qassem, J. Ng, Augmented immersive reality (AIR) for improved learning performance: a quantitative evaluation, IEEE Trans. Learn. Technol. 13 (2020) 283–296, https://doi.org/10.1109/TLT.2019.2937525.
[3] E. Schiza, M. Matsangidou, K. Neokleous, C.S. Pattichis, Virtual reality applications for neurological disease: a review, Front. Robot. AI 6 (2019) 100, https://doi.org/10.3389/frobt.2019.00100.
[4] S. Słyk, M.Z. Zarzycki, A. Kocwa-Karnaś, I. Domitrz, Virtual reality in the diagnostics and therapy of neurological diseases, Expert Rev. Med. Devices 16 (2019) 1035–1040, https://doi.org/10.1080/17434440.2019.1693892.
[5] Y. Liu, W. Tan, C. Chen, C. Liu, J. Yang, Y. Zhang, A review of the application of virtual reality technology in the diagnosis and treatment of cognitive impairment, Front. Aging Neurosci. 11 (2019) 280, https://doi.org/10.3389/fnagi.2019.00280.
[6] G. Bresson, Z. Alsayed, L. Yu, S. Glaser, Simultaneous localization and mapping: a survey of current trends in autonomous driving, IEEE Trans. Intell. Veh. 2 (2017) 194–220, https://doi.org/10.1109/TIV.2017.2749181.
[7] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, J.J. Leonard, Past, present, and future of simultaneous localization and mapping: toward the robust-perception age, IEEE Trans. Robot. 32 (2016) 1309–1332, https://doi.org/10.1109/TRO.2016.2624754.
[8] U. Frese, A discussion of simultaneous localization and mapping, Auton. Robot. 20 (2006) 25–42, https://doi.org/10.1007/s10514-006-5735-x.
[9] R. Chatila, J. Laumond, Position referencing and consistent world modeling for mobile robots, in: Proceedings. 1985 IEEE International Conference on Robotics and Automation, 1985, pp. 138–145, https://doi.org/10.1109/ROBOT.1985.1087373.
[10] T. Taketomi, H. Uchiyama, S. Ikeda, Visual SLAM algorithms: a survey from 2010 to 2016, IPSJ Trans. Comput. Vis. Appl. 9 (2017) 16, https://doi.org/10.1186/s41074-017-0027-2.
[11] F. Yu, J. Shang, Y. Hu, M. Milford, NeuroSLAM: a brain-inspired SLAM system for 3D environments, Biol. Cybern. 113 (2019) 515–545, https://doi.org/10.1007/s00422-019-00806-9.
[12] M.J. Milford, G.F. Wyeth, D. Prasser, RatSLAM: a hippocampal model for simultaneous localization and mapping, in: IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 1, 2004, pp. 403–408, https://doi.org/10.1109/ROBOT.2004.1307183.
[13] M. Chancán, L. Hernandez-Nunez, A. Narendra, A.B. Barron, M. Milford, A hybrid compact neural architecture for visual place recognition, IEEE Robot Autom. Lett. 5 (2020) 993–1000, https://doi.org/10.1109/LRA.2020.2967324.
[14] S.-C. Zhou, R. Yan, J.-X. Li, Y.-K. Chen, H. Tang, A brain-inspired SLAM system based on ORB features, Int. J. Autom. Comput. 14 (2017) 564–575, https://doi.org/10.1007/s11633-017-1090-y.
[15] G. Bresson, T. Féraud, R. Aufrère, P. Checchin, R. Chapuis, Real-time monocular SLAM with low memory requirements, IEEE Trans. Intell. Transp. Syst. 16 (2015) 1827–1839, https://doi.org/10.1109/TITS.2014.2376780.
[16] F. Munoz-Montoya, M.-C. Juan, M. Mendez-Lopez, R. Molla, F. Abad, C. Fidalgo, SLAM-based augmented reality for the assessment of short-term spatial memory. A comparative study of visual versus tactile stimuli, PLoS ONE 16 (2021) e0245976.
[17] L.E.H. van Dokkum, T. Ward, I. Laffont, Brain computer interfaces for neurorehabilitation – its current status as a rehabilitation strategy post-stroke, Ann. Phys. Rehabil. Med. 58 (2015) 3–8, https://doi.org/10.1016/j.rehab.2014.09.016.
[18] F. Schirmann, The wondrous eyes of a new technology"—a history of the early electroencephalography (EEG) of psychopathy, delinquency, and immorality, Front. Hum. Neurosci. 8 (2014) 232, https://doi.org/10.3389/fnhum.2014.00232.
[19] T. Ma, H. Li, L. Deng, H. Yang, X. Lv, P. Li, F. Li, R. Zhang, T. Liu, D. Yao, P. Xu, The hybrid BCI system for movement control by combining motor imagery and moving onset visual evoked potential, J. Neural Eng. 14 (2017) 26015, https://doi.org/10.1088/1741-2552/aa5d5f.
[20] D. Yao, Y. Qin, S. Hu, L. Dong, M.L. Bringas Vega, P.A. Valdés Sosa, Which reference should we use for EEG and ERP practice? Brain Topogr. 32 (2019) 530–549, https://doi.org/10.1007/s10548-019-00707-x.
[21] S. Milanizadeh, J. Safaie, EOG-based HCI system for quadcopter navigation, IEEE Trans. Instrum. Meas. 69 (2020) 8992–8999, https://doi.org/10.1109/TIM.2020.3001411.
[22] J.L. Collinger, B. Wodlinger, J.E. Downey, W. Wang, E.C. Tyler-Kabara, D.J. Weber, A.J.C. McMorland, M. Velliste, M.L. Boninger, A.B. Schwartz, High-performance neuroprosthetic control by an individual with tetraplegia, Lancet 381 (2013) 557–564, https://doi.org/10.1016/S0140-6736(12)61816-9.

[23] K.J. Miller, D. Hermes, N.P. Staff, The current state of electrocorticography-based brain-computer interfaces, Neurosurg. Focus 49 (2020) E2, https://doi.org/10.3171/2020.4.FOCUS20185.

[24] S. Chakrabarti, H.M. Sandberg, J.S. Brumberg, D.J. Krusienski, Progress in speech decoding from the electrocorticogram, Biomed. Eng. Lett. 5 (2015) 10–21, https://doi.org/10.1007/s13534-015-0175-1.

[25] A. Kübler, The history of BCI: from a vision for the future to real support for personhood in people with locked-in syndrome, Neuroethics 13 (2020) 1–18, https://doi.org/10.1007/s12152-019-09409-4.

[26] E.C. Leuthardt, K.J. Miller, G. Schalk, R.P.N. Rao, J.G. Ojemann, Electrocorticography-based brain computer interface–the Seattle experience, IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2006) 194–198, https://doi.org/10.1109/TNSRE.2006.875536.

[27] M. Bamdad, H. Zarshenas, M.A. Auais, Application of BCI systems in neurorehabilitation: a scoping review, Disabil. Rehabil. Assist. Technol. 10 (2015) 355–364, https://doi.org/10.3109/17483107.2014.961569.

[28] A. Dubey, S. Ray, Cortical electrocorticogram (ECoG) is a local signal, J. Neurosci. 39 (2019) 4299–4311, https://doi.org/10.1523/JNEUROSCI.2917-18.2019.

[29] M.A. Lebedev, M.A.L. Nicolelis, Brain–machine interfaces: past, present and future, Trends Neurosci. 29 (2006) 536–546, https://doi.org/10.1016/j.tins.2006.07.004.

[30] S. Kolhe, D. Khemani, C. Bhatt, N. Dubey, Automation of appliances using electroencephalography: virtual reality, augmented reality, artificial intelligence, internet of things, robotics, industry 4.0, in: (2018): pp. 223–243. 10.1002/9781119509875.ch17.

[31] N.R.B. Martins, A. Angelica, K. Chakravarthy, Y. Svidinenko, F.J. Boehm, I. Opris, M.A. Lebedev, M. Swan, S.A. Garan, J.V. Rosenfeld, T. Hogg, R.A. Freitas, Human brain/cloud interface, Front. Neurosci. 13 (2019) 112, https://doi.org/10.3389/fnins.2019.00112.

[32] J.L. Collinger, M.A. Kryger, R. Barbara, T. Betler, K. Bowsher, E.H.P. Brown, S.T. Clanton, A.D. Degenhart, S.T. Foldes, R.A. Gaunt, F.E. Gyulai, E.A. Harchick, D. Harrington, J.B. Helder, T. Hemmes, M.S. Johannes, K.D. Katyal, G.S.F. Ling, A.J.C. McMorland, K. Palko, M.P. Para, J. Scheuermann, A.B. Schwartz, E. R. Skidmore, F. Solzbacher, A.V. Srikameswaran, D.P. Swanson, S. Swetz, E.C. Tyler-Kabara, M. Velliste, W. Wang, D.J. Weber, B. Wodlinger, M.L. Boninger, Collaborative approach in the development of high-performance brain–computer interfaces for a neuroprosthetic arm: translation from animal models to human control, Clin. Transl. Sci. 7 (2014) 52–59, https://doi.org/10.1111/cts.12086.

[33] C.E. Bouton, A. Shaikhouni, N.V. Annetta, M.A. Bockbrader, D.A. Friedenberg, D.M. Nielson, G. Sharma, P.B. Sederberg, B.C. Glenn, W.J. Mysiw, A.G. Morgan, M. Deogaonkar, A.R. Rezai, Restoring cortical control of functional movement in a human with quadriplegia, Nature 533 (2016) 247–250, https://doi.org/10.1038/nature17435.

[34] P.D. Ganzer, S.C. Colachis, M.A. Schwemmer, D.A. Friedenberg, C.F. Dunlap, C.E. Swiftney, A.F. Jacobowitz, D.J. Weber, M.A. Bockbrader, G. Sharma, Restoring the sense of touch using a sensorimotor demultiplexing neural interface, Cell 181 (2020) 763–773, https://doi.org/10.1016/j.cell.2020.03.054, e12.

[35] D. Kapgate, Future of EEG based hybrid visual brain computer interface systems in rehabilitation of people with neurological disorders, Int. Res. J. Adv. Sci. Hub 2 (2020) 15–20, https://doi.org/10.47392/irjash.2020.31.

[36] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, J.R. Wolpaw, BCI2000: a general-purpose brain-computer interface (BCI) system, IEEE Trans. Biomed. Eng. 51 (2004) 1034–1043, https://doi.org/10.1109/TBME.2004.827072.

[37] J. Kim, J. Lee, C. Han, K. Park, An instant donning multi-channel EEG headset (with comb-shaped dry electrodes) and BCI applications, Sensors 19 (2019), https://doi.org/10.3390/s19071537.

[38] M.A.H. Hasan, M.U. Khan, D. Mishra, A computationally efficient method for hybrid EEG-fNIRS BCI based on the Pearson correlation, Biomed. Res. Int. 2020 (2020) 1838140, https://doi.org/10.1155/2020/1838140.

[39] J.Del R. Millán, P.W. Ferrez, F. Galán, E. Lew, R. Chavarriaga, Non-invasive brain-machine interaction, Intern. J. Pattern Recognit. Artif. Intell. 22 (2008) 959–972, https://doi.org/10.1142/S0218001408006600.

[40] F. Cincotti, D. Mattia, F. Aloise, S. Bufalari, G. Schalk, G. Oriolo, A. Cherubini, M.G. Marciani, F. Babiloni, Non-invasive brain–computer interface system: towards its application as assistive technology, Brain Res. Bull. 75 (2008) 796–803, https://doi.org/10.1016/j.brainresbull.2008.01.007.

[41] A. Ravi, N.H. Beni, J. Manuel, N. Jiang, Comparing user-dependent and user-independent training of CNN for SSVEP BCI, J. Neural Eng. 17 (2020) 026028, https://doi.org/10.1088/1741-2552/ab6a67.

[42] P. Schembri, M. Pelc, J. Ma, The effect that auxiliary taxonomized auditory distractions have on a p300 speller while utilising low fidelity equipment, in: 2019 11th Computer Science and Electronic Engineering (CEEC), 2019, pp. 118–123, https://doi.org/10.1109/CEEC47804.2019.8974315.

[43] B.Z. Allison, A. Kübler, J. Jin, 30+ years of P300 brain–computer interfaces, Psychophysiology 57 (2020) e13569, https://doi.org/10.1111/psyp.13569.

[44] I.A. Fouad, F.E.-Z.M. Labib, M.S. Mabrouk, A.A. Sharawy, A.Y. Sayed, Improving the performance of P300 BCI system using different methods, Netw. Model. Anal. Health Inform. Bioinform. 9 (2020) 64, https://doi.org/10.1007/s13721-020-00268-1.

[45] B. Peters, S. Bedrick, S. Dudy, B. Eddy, M. Higger, M. Kinsella, D. McLaughlin, T. Memmott, B. Oken, F. Quivira, S. Spaulding, D. Erdogmus, M. Fried-Oken, SSVEP BCI and eye tracking use by individuals with late-stage ALS and visual impairments, Front. Hum. Neurosci. 14 (2020) 595890, https://doi.org/10.3389/fnhum.2020.595890.

[46] H.-J. Hwang, J.-H. Lim, Y.-J. Jung, H. Choi, S.W. Lee, C.-H. Im, Development of an SSVEP-based BCI spelling system adopting a QWERTY-style LED keyboard, J. Neurosci. Methods 208 (2012) 59–65, https://doi.org/10.1016/j.jneumeth.2012.04.011.

[47] G.A.M. Vasiljevic, L.C. de Miranda, Brain–computer interface games based on consumer-grade EEG devices: a systematic literature review, Int. J. Hum. Comput. Interact. 36 (2020) 105–142, https://doi.org/10.1080/10447318.2019.1612213.

[48] Y.M. Chi, Y. Wang, Y. Wang, C. Maier, T. Jung, G. Cauwenberghs, Dry and noncontact EEG sensors for mobile brain–computer interfaces, IEEE Trans. Neural Syst. Rehabil. Eng. 20 (2012) 228–235, https://doi.org/10.1109/TNSRE.2011.2174652.

[49] J. LaRocco, M.D. Le, D.-G. Paeng, A systemic review of available low-cost EEG headsets used for drowsiness detection, Front. Neuroinform. 14 (2020) 553352, https://doi.org/10.3389/fninf.2020.553352.

[50] P. de Lissa, S. Sörensen, N. Badcock, J. Thie, G. McArthur, Measuring the face-sensitive N170 with a gaming EEG system: a validation study, J. Neurosci. Methods 253 (2015) 47–54, https://doi.org/10.1016/j.jneumeth.2015.05.025.

[51] A. Aghaei-Lasboo, K. Inoyama, A.S. Fogarty, J. Kuo, K.J. Meador, J.J. Walter, S.T. Le, K.D. Graber, B. Razavi, R.S. Fisher, Tripolar concentric EEG electrodes reduce noise, Clin. Neurophysiol. 131 (2020) 193–198, https://doi.org/10.1016/j.clinph.2019.10.022.

[52] X. Liu, O. Makeyev, W. Besio, Improved spatial resolution of electroencephalogram using tripolar concentric ring electrode sensors, J. Sens. 2020 (2020) 6269394, https://doi.org/10.1155/2020/6269394.

[53] A. Kawala-Sterniuk, N. Browarska, A. Al-Bakri, M. Pelc, J. Zygarlicki, M. Sidikova, R. Martinek, E.J. Gorzelanczyk, Summary of over fifty years with brain-computer interfaces—a review, Brain Sci. 11 (2021), https://doi.org/10.3390/brainsci11010043.

[54] M.R. Mowla, R.I. Cano, K.J. Dhuyvetter, D.E. Thompson, Affective brain-computer interfaces: choosing a meaningful performance measuring metric, Comput. Biol. Med. 126 (2020) 104001, https://doi.org/10.1016/j.compbiomed.2020.104001.

[55] J.N. Mak, J.R. Wolpaw, Clinical applications of brain-computer interfaces: current state and future prospects, IEEE Rev. Biomed. Eng. 2 (2009) 187–199, https://doi.org/10.1109/RBME.2009.2035356.

[56] M. Sebastián-Romagosa, W. Cho, R. Ortner, N. Murovec, T. Von Oertzen, K. Kamada, B.Z. Allison, C. Guger, Brain computer interface treatment for motor rehabilitation of upper extremity of stroke patients-a feasibility study, Front. Neurosci. 14 (2020) 591435, https://doi.org/10.3389/fnins.2020.591435.

[57] G.M. Wojcik, J. Masiak, A. Kawiak, P. Schneider, L. Kwasniewicz, N. Polak, A. Gajos-Balinska, New protocol for quantitative analysis of brain cortex electroencephalographic activity in patients with psychiatric disorders, Front. Neuroinform. 12 (2018) 27, https://doi.org/10.3389/fninf.2018.00027.

[58] J. Nirme, B. Rubio, A. Duff, E. Duarte, S. Rodriguez, A. Cuxart, P.F.M.J. Verschure, At home motor rehabilitation in the chronic phase of stroke using the rehabilitation gaming system. Biosystems and Biorobotics, Springer International Publishing, 2013, pp. 931–935, https://doi.org/10.1007/978-3-642-34546-3_151.

[59] M. Cameirão, S. Bermúdez i Badia, E. Duarte, P. Verschure, The rehabilitation gaming system: a review, Stud. Health Technol. Inform. 145 (2009) 65–83, https://doi.org/10.3233/978-1-60750-018-6-65.

[60] M.S. Cameirão, S.B. i. Badia, E.D. Oller, P.F.M.J. Verschure, Neurorehabilitation using the virtual reality based rehabilitation gaming system: methodology, design, psychometrics, usability and validation, J. Neuroeng. Rehabil. 7 (2010) 48, https://doi.org/10.1186/1743-0003-7-48.
[61] M.S. Cameirao, S.B.i. Badia, L. Zimmerli, E.D. Oller, P.F.M.J. Verschure, The rehabilitation gaming system: a virtual reality based system for the evaluation and rehabilitation of motor deficits, in: 2007 Virtual Rehabilitation, 2007, pp. 29–33, https://doi.org/10.1109/ICVR.2007.4362125.
[62] T.T. James, A. Brammatha, Effect of gaming assisted visual feedback on functional standing balance among acute hemiparetic stroke patients, Indian J. Physiother. Occup. Ther 11 (2017) 151–155, https://doi.org/10.5958/0973-5674.2017.00137.X.
[63] B.R. Ballester, M. Maier, D.A. Domingo, A. Aguilar, A. Mura, L.T. Pareja, M.F.G. Esteve, P.F.M.J. Verschure, Adaptive VR-based rehabilitation to prevent deterioration in adults with cerebral palsy, in: 2019 International Conference on Virtual Rehabilitation (ICVR), 2019, pp. 1–7, https://doi.org/10.1109/ICVR46560.2019.8994754.
[64] F. Covaciu, A. Pisla, C. Vaida, B. Gherman, D. Pisla, Development of a virtual reality simulator for a lower limb rehabilitation robot, in: 2020 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR), 2020, pp. 1–6, https://doi.org/10.1109/AQTR49680.2020.9129981.
[65] B. Aruanno, G. Caruso, M. Rossini, F. Molteni, M.C.E. Espinoza, M. Covarrubias, Virtual and augmented reality platform for cognitive tele-rehabilitation based system. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Science and Business Media Deutschland GmbH, 2020, pp. 130–137, https://doi.org/10.1007/978-3-030-58796-3_17.
[66] M.I. Daoud, A. Alhusseini, M.Z. Ali, R. Alazrai, A game-based rehabilitation system for upper-limb cerebral palsy: a feasibility study, Sensors 20 (2020), https://doi.org/10.3390/s20082416.
[67] G. Fuertes Muñoz, R.A. Mollineda, J. Gallardo Casero, F. Pla, A RGBD-based interactive system for gaming-driven rehabilitation of upper limbs, Sensors 19 (2019), https://doi.org/10.3390/s19163478.
[68] M. Välimäki, K. Mishina, J.K. Kaakinen, S.K. Holm, J. Vahlo, M. Kirjonen, V. Pekurinen, O. Tenovuo, J. Korkeila, H. Hämäläinen, J. Sarajuuri, P. Rantanen, T. Orenius, A. Koponen, Digital gaming for improving the functioning of people with traumatic brain injury: randomized clinical feasibility study, J. Med. Internet Res. 20 (2018) e77, https://doi.org/10.2196/jmir.7618.
[69] V. Jayasree-Krishnan, S.R. Bethi, S. Kumar, R.K.R. Jayanthi, S. Ghosh, P. Raghavan, V. Kapila, RehabFork: an interactive game-assisted upper limb stroke rehabilitation system, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 5757–5760, https://doi.org/10.1109/EMBC44109.2020.9176168.
[70] S. Tresser, T. Kuflik, I. Levin, P.L.T. Weiss, Validation of a novel personalized therapeutic virtual gaming system, in: 2019 International Conference on Virtual Rehabilitation (ICVR), 2019, pp. 1–6, https://doi.org/10.1109/ICVR46560.2019.8994440.
[71] D. Dall'Alba, I. Dimbwadyo, S. Piazza, E. Magnabosco, G. Menegozzo, P. Fiorini, ViTAS gaming suite: virtual therapy against stroke. Biosystems and Biorobotics, Springer International Publishing, 2017, pp. 407–411, https://doi.org/10.1007/978-3-319-46669-9_68.
[72] S. Thomas, L. Fazakarley, P.W. Thomas, S. Collyer, S. Brenton, S. Perring, R. Scott, F. Thomas, C. Thomas, K. Jones, J. Hickson, C. Hillier, Mii-vitaliSe: a pilot randomised controlled trial of a home gaming system (Nintendo Wii) to increase activity levels, vitality and well-being in people with multiple sclerosis, BMJ Open 7 (2017) e016966, https://doi.org/10.1136/bmjopen-2017-016966.
[73] K. Cleary, R. Monfaredi, T. Salvador, H.F. Talari, C. Coley, S. Kovelman, J. Belschner, S. Alyamani, M. Schladen, S.H. Evans, Pedbothome: robotically-assisted ankle rehabilitation system for children with cerebral palsy*, in: 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR), 2019, pp. 13–20, https://doi.org/10.1109/ICORR.2019.8779468.
[74] H. Tannous, D. Istrate, A. Perrochon, J.-C. Daviet, A. Benlarbi-Delai, J. Sarrazin, M.-C.H.B. Tho, T.T. Dao, GAMEREHAB@HOME: a new engineering system using serious game and multisensor fusion for functional rehabilitation at home, IEEE Trans. Games 13 (2021) 89–98, https://doi.org/10.1109/TG.2019.2963108.
[75] S.S. Esfahlani, T. Thompson, A.D. Parsa, I. Brown, S. Cirstea, ReHabgame: a non-immersive virtual reality rehabilitation system with applications in neuroscience, Heliyon 4 (2018) e00526, https://doi.org/10.1016/j.heliyon.2018.e00526.
[76] M. Kefalas, M. Koch, V. Geraedts, H. Wang, M. Tannemaat, T. Bäck, Automated machine learning for the classification of normal and abnormal electromyography data, in: 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 1176–1185, https://doi.org/10.1109/BigData50022.2020.9377780.
[77] E. Stlberg, B. Falck, The role of electromyography in neurology, Electroencephalogr. Clin. Neurophysiol. 103 (1997) 579–598, https://doi.org/10.1016/S0013-4694(97)00138-7.
[78] K.R. Mills, The basics of electromyography, J. Neurol., Neurosurg. Psychiatry 76 (2005), https://doi.org/10.1136/jnnp.2005.069211 ii32 LP-ii35.
[79] E. Scheme, K. Englehart, Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use, J. Rehabil. Res. Dev. 48 (2011) 643–659, https://doi.org/10.1682/jrrd.2010.09.0177.
[80] A. Phinyomark, P. Phukpattaranont, C. Limsakul, A review of control methods for electric power wheelchairs based on electromyography signals with special emphasis on pattern recognition, IETE Tech. Rev. 28 (2011) 316–326, https://doi.org/10.4103/0256-4602.83552.
[81] T.S. Saponas, D.S. Tan, D. Morris, R. Balakrishnan, J. Turner, J.A. Landay, Enabling always-available input with muscle-computer interfaces, in: Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology, New York, NY, USA, Association for Computing Machinery, 2009, pp. 167–176, https://doi.org/10.1145/1622176.1622208.
[82] Z. Lei, An upper limb movement estimation from electromyography by using BP neural network, Biomed. Signal Process. Control 49 (2019) 434–439, https://doi.org/10.1016/j.bspc.2018.12.020.
[83] O. Ulkir, G. Gokmen, E. Kaplanoglu, Emg signal classification using fuzzy logic, Balkan J. Electr. Comput. Eng. 5 (2017) 97–101, https://doi.org/10.17694/bajece.337941.
[84] M.A. Ozdemir, D.H. Kisa, O. Guren, A. Onan, A. Akan, EMG based hand gesture recognition using deep learning, in: 2020 Medical Technologies Congress (TIPTEKNO), 2020, pp. 1–4, https://doi.org/10.1109/TIPTEKNO50054.2020.9299264.
[85] N. Anil, S.H. Sreeletha, EMG based gesture recognition using machine learning, in: 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 1560–1564, https://doi.org/10.1109/ICCONS.2018.8662987.
[86] D. Huang, B. Chen, Surface EMG decoding for hand gestures based on spectrogram and CNN-LSTM, in: 2019 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI), 2019, pp. 123–126, https://doi.org/10.1109/CCHI.2019.8901936.
[87] P. Tsinganos, B. Cornelis, C. Jan, B. Jansen, A. Skodras, Deep learning in EMG-based gesture recognition, (2018). https://doi.org/10.5220/0006960201070114.
[88] R.J. Duarte-Moreira, K.V.-F. Castro, C. Luz-Santos, J.V.P. Martins, K.N. Sá, A.F. Baptista, Electromyographic biofeedback in motor function recovery after peripheral nerve injury: an integrative review of the literature, Appl. Psychophysiol .Biofeedback 43 (2018) 247–257, https://doi.org/10.1007/s10484-018-9403-7.
[89] H. Woodford, C. Price, EMG biofeedback for the recovery of motor function after stroke, Cochrane Database Syst. Rev. 2007 (2007), https://doi.org/10.1002/14651858.CD004585.pub2. CD004585–CD004585.
[90] A.B. Gámez, J.J. Hernandez Morante, J.L. Martínez Gil, F. Esparza, C.M. Martínez, The effect of surface electromyography biofeedback on the activity of extensor and dorsiflexor muscles in elderly adults: a randomized trial, Sci. Rep. 9 (2019) 13153, https://doi.org/10.1038/s41598-019-49720-x.
[91] B. Barth, K. Mayer, U. Strehl, A.J. Fallgatter, A.-C. Ehlis, EMG biofeedback training in adult attention-deficit/hyperactivity disorder: an active (control) training? Behav. Brain Res. 329 (2017) 58–66, https://doi.org/10.1016/j.bbr.2017.04.021.
[92] O.M. Giggins, U.M. Persson, B. Caulfield, Biofeedback in rehabilitation, J. Neuroeng. Rehabil. 10 (2013) 60, https://doi.org/10.1186/1743-0003-10-60.
[93] T.W. Boonstra, L. Faes, J.N. Kerkman, D. Marinazzo, Information decomposition of multichannel EMG to map functional interactions in the distributed motor system, Neuroimage 202 (2019) 116093, https://doi.org/10.1016/j.neuroimage.2019.116093.
[94] P. Bugalho, F. Ladeira, R. Barbosa, J.P. Marto, C. Borbinha, L. da Conceição, M. Salavisa, M. Saraiva, B. Meira, M. Fernandes, Polysomnographic predictors of sleep, motor and cognitive dysfunction progression in Parkinson's disease: a longitudinal study, Sleep Med. 77 (2021) 205–208, https://doi.org/10.1016/j.sleep.2020.06.020.

[95] C. Schmit, C. Hausswirth, Y. Le Meur, R. Duffield, Cognitive functioning and heat strain: performance responses and protective strategies, Sports Med. 47 (2017) 1289–1302, https://doi.org/10.1007/s40279-016-0657-z.

[96] S.A. Acuña, C.A. Francis, J.R. Franz, D.G. Thelen, The effects of cognitive load and optical flow on antagonist leg muscle coactivation during walking for young and older adults, J. Electromyogr. Kinesiol. 44 (2019) 8–14, https://doi.org/10.1016/j.jelekin.2018.11.003.

[97] Z. Zhang, Y. Gao, J. Wang, Effects of vision and cognitive load on anticipatory and compensatory postural control, Hum. Mov. Sci. 64 (2019) 398–408, https://doi.org/10.1016/j.humov.2019.02.011.

[98] F. Li, S. Yin, P. Feng, N. Hu, C. Ding, A. Chen, The cognitive up- and down-regulation of positive emotion: evidence from behavior, electrophysiology, and neuroimaging, Biol. Psychol. 136 (2018) 57–66, https://doi.org/10.1016/j.biopsycho.2018.05.013.

[99] Y. Liu, J. Zhang, S.P. Lam, M.W.M. Yu, S.X. Li, J. Zhou, N.Y. Chan, J. Wang, H. Feng, A. Chan, V. Mok, Y.K. Wing, Electromyography activity level in rapid eye movement sleep predicts neurodegenerative diseases in idiopathic rapid eye movement sleep behavior disorder: a 5-year longitudinal study, Sleep Med. 56 (2019) 128–134, https://doi.org/10.1016/j.sleep.2019.01.018.

[100] C. Loconsole, G.F. Trotta, A. Brunetti, J. Trotta, A. Schiavone, S.I. Tatò, G. Losavio, V. Bevilacqua, Computer vision and EMG-based handwriting analysis for classification in Parkinson's disease. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 2017, pp. 493–503, https://doi.org/10.1007/978-3-319-63312-1_43.

[101] M. Skelly, A. Salameh, J. McCabe, S. Pundik, MEP-ART: a system for real-time feedback and analysis of transcranial magnetic stimulation motor evoked potentials, Brain Stimul. 13 (2020) 1614–1616, https://doi.org/10.1016/j.brs.2020.09.012.

[102] N. Jackson, I. Greenhouse, VETA: an open-source matlab-based toolbox for the collection and analysis of electromyography combined with transcranial magnetic stimulation, Front. Neurosci. 13 (2019) 975, https://doi.org/10.3389/fnins.2019.00975.

[103] S. Phadikar, N. Sinha, R. Ghosh, Automatic eyeblink artifact removal from EEG signal using wavelet transform with heuristically optimized threshold, IEEE J. Biomed. Health Inform. 25 (2021) 475–484, https://doi.org/10.1109/JBHI.2020.2995235.

[104] A. Biasiucci, B. Franceschiello, M.M. Murray, Electroencephalography, Curr. Biol. 29 (2019) R80–R85, https://doi.org/10.1016/j.cub.2018.11.052.

[105] J. Parvizi, S. Kastner, Promises and limitations of human intracranial electroencephalography, Nat. Neurosci. 21 (2018) 474–483, https://doi.org/10.1038/s41593-018-0108-2.

[106] L. Duan, J. Li, H. Ji, Z. Pang, X. Zheng, R. Lu, M. Li, J. Zhuang, Zero-shot learning for EEG classification in motor imagery-based BCI system, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (2020) 2411–2419, https://doi.org/10.1109/TNSRE.2020.3027004.

[107] H. Namazi, E. Aghasian, T.S. Ala, Fractal-based classification of electroencephalography (EEG) signals in healthy adolescents and adolescents with symptoms of schizophrenia, Technol. Health Care 27 (2019) 233–241, https://doi.org/10.3233/THC-181497.

[108] S. Nobukawa, T. Yamanishi, S. Kasakawa, H. Nishimura, M. Kikuchi, T. Takahashi, Classification methods based on complexity and synchronization of electroencephalography signals in Alzheimer's disease, Front. Psychiatry 11 (2020) 255, https://doi.org/10.3389/fpsyt.2020.00255.

[109] N. Betrouni, A. Delval, L. Chaton, L. Defebvre, A. Duits, A. Moonen, A.F.G. Leentjens, K. Dujardin, Electroencephalography-based machine learning for cognitive profiling in Parkinson's disease: preliminary results, Mov. Disord. 34 (2019) 210–217, https://doi.org/10.1002/mds.27528.

[110] B. Freund, J.C. Probasco, E.K. Ritzl, Seizure incidence in the acute postneurosurgical period diagnosed using continuous electroencephalography, J. Neurosurg. JNS. 130 (2018) 1203–1209, https://doi.org/10.3171/2018.1.JNS171466.

[111] R. Cassani, T.H. Falk, F.J. Fraga, M. Cecchi, D.K. Moore, R. Anghinah, Towards automated electroencephalography-based Alzheimer's disease diagnosis using portable low-density devices, Biomed. Signal Process Control 33 (2017) 261–271, https://doi.org/10.1016/j.bspc.2016.12.009.

[112] B. Min, M. Kim, J. Lee, J.-I. Byun, K. Chu, K.-Y. Jung, S.K. Lee, J.S. Kwon, Prediction of individual responses to electroconvulsive therapy in patients with schizophrenia: machine learning analysis of resting-state electroencephalography, Schizophr. Res. 216 (2020) 147–153, https://doi.org/10.1016/j.schres.2019.12.012.

[113] S. Ibrahim, R. Djemal, A. Alsuwailem, Electroencephalography (EEG) signal processing for epilepsy and autism spectrum disorder diagnosis, Biocybern. Biomed. Eng. 38 (2018) 16–26, https://doi.org/10.1016/j.bbe.2017.08.006.

[114] U. Seneviratne, M.J. Cook, W.J. D'Souza, Electroencephalography in the diagnosis of genetic generalized epilepsy syndromes, Front. Neurol. 8 (2017) 499, https://doi.org/10.3389/fneur.2017.00499.

[115] J. Hui, S. Tremblay, Z.J. Daskalakis, The current and future potential of transcranial magnetic stimulation with electroencephalography in psychiatry, Clin. Pharmacol. Ther. 106 (2019) 734–746, https://doi.org/10.1002/cpt.1541.

[116] R.T. Toll, W. Wu, S. Naparstek, Y. Zhang, M. Narayan, B. Patenaude, C. De Los Angeles, K. Sarhadi, N. Anicetti, P. Longwell, E. Shpigel, R. Wright, J. Newman, B. Gonzalez, R. Hart, S. Mann, D. Abu-Amara, K. Sarhadi, C. Cornelssen, C. Marmar, A. Etkin, An electroencephalography connectomic profile of posttraumatic stress disorder, Am. J. Psychiatry 177 (2020) 233–243, https://doi.org/10.1176/appi.ajp.2019.18080911.

[117] A. Emami, N. Kunii, T. Matsuo, T. Shinozaki, K. Kawai, H. Takahashi, Seizure detection by convolutional neural network-based analysis of scalp electroencephalography plot images, Neuroimage Clin. 22 (2019) 101684, https://doi.org/10.1016/j.nicl.2019.101684.

[118] J. Naim-Feil, M. Rubinson, D. Freche, A. Grinshpoon, A. Peled, E. Moses, N. Levit-Binnun, Altered brain network dynamics in schizophrenia: a cognitive electroencephalography study, Biol. Psychiatry Cogn. Neurosci. Neuroimaging 3 (2018) 88–98, https://doi.org/10.1016/j.bpsc.2017.03.017.

[119] A.A. Bhurane, S. Dhok, M. Sharma, R. Yuvaraj, M. Murugappan, U.R. Acharya, Diagnosis of Parkinson's disease from electroencephalography signals using linear and self-similarity features, Expert Syst. n/a (2019) e12472, https://doi.org/10.1111/exsy.12472.

[120] B. Oltu, M.F. Akşahin, S. Kibaroğlu, A novel electroencephalography based approach for Alzheimer's disease and mild cognitive impairment detection, Biomed. Signal Process. Control 63 (2021) 102223, https://doi.org/10.1016/j.bspc.2020.102223.

[121] S. Lalgudi Ganesan, C.P. Stewart, E.G. Atenafu, R. Sharma, A.-M. Guerguerian, J.S. Hutchison, C.D. Hahn, Seizure identification by critical care providers using quantitative electroencephalography, Crit. Care Med. 46 (2018).

[122] M.J. Schiller, Quantitative electroencephalography in guiding treatment of major depression, Front. Psychiatry 9 (2019) 779, https://doi.org/10.3389/fpsyt.2018.00779.

[123] P. Bourdillon, M. Cucherat, J. Isnard, K. Ostrowsky-Coste, H. Catenoix, M. Guénot, S. Rheims, Stereo-electroencephalography-guided radiofrequency thermocoagulation in patients with focal epilepsy: a systematic review and meta-analysis, Epilepsia 59 (2018) 2296–2304, https://doi.org/10.1111/epi.14584.

[124] A. Appriou, A. Cichocki, F. Lotte, Modern machine-learning algorithms: for classifying cognitive and affective states from electroencephalography signals, IEEE Syst. Man Cybern. Mag. 6 (2020) 29–38, https://doi.org/10.1109/MSMC.2020.2968638.

[125] N. Jackson, S.R. Cole, B. Voytek, N.C. Swann, Characteristics of waveform shape in Parkinson's disease detected with scalp electroencephalography, eNeuro 6 (2019), https://doi.org/10.1523/ENEURO.0151-19.2019. ENEURO.0151-19.2019.

[126] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T.H. Falk, J. Faubert, Deep learning-based electroencephalography analysis: a systematic review, J. Neural Eng. 16 (2019) 51001, https://doi.org/10.1088/1741-2552/ab260c.

[127] C. Tan, F. Sun, W. Zhang, J. Chen, C. Liu, Multimodal classification with deep convolutional-recurrent neural networks for electroencephalography. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 2017, pp. 767–776, https://doi.org/10.1007/978-3-319-70096-0_78.

[128] Y.-H. Kwon, S.-B. Shin, S.-D. Kim, Electroencephalography based fusion two-dimensional (2D)-convolution neural networks (CNN) model for emotion recognition system, Sensors 18 (2018), https://doi.org/10.3390/s18051383.

[129] F. Movahedi, J.L. Coyle, E. Sejdić, Deep belief networks for electroencephalography: a review of recent contributions and future outlooks, IEEE J. Biomed. Health Inform. 22 (2018) 642–652, https://doi.org/10.1109/JBHI.2017.2727218.

[130] R. Maldonado, S.M. Harabagiu, Active deep learning for the identification of concepts and relations in electroencephalography reports, J. Biomed. Inform. 98 (2019) 103265, https://doi.org/10.1016/j.jbi.2019.103265.

[131] N. Jaworska, S. de la Salle, M.-H. Ibrahim, P. Blier, V. Knott, Leveraging machine learning approaches for predicting antidepressant treatment response using electroencephalography (EEG) and clinical data, Front. Psychiatry 9 (2019) 768, https://doi.org/10.3389/fpsyt.2018.00768.

[132] T. de Taillez, B. Kollmeier, B.T. Meyer, Machine learning for decoding listeners' attention from electroencephalography evoked by continuous speech, Eur. J. Neurosci. 51 (2020) 1234–1241, https://doi.org/10.1111/ejn.13790.
[133] M. Ploner, E.S. May, Electroencephalography and magnetoencephalography in pain research—current state and future perspectives, Pain 159 (2018).
[134] M.C. Piastra, A. Nüßing, J. Vorwerk, M. Clerc, C. Engwer, C.H. Wolters, A comprehensive study on electroencephalography and magnetoencephalography sensitivity to cortical and subcortical sources, Hum. Brain Mapp. 42 (2021) 978–992, https://doi.org/10.1002/hbm.25272.
[135] N. Ding, L. Melloni, A. Yang, Y. Wang, W. Zhang, D. Poeppel, Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG), Front. Hum. Neurosci. 11 (2017) 481, https://doi.org/10.3389/fnhum.2017.00481.
[136] Q. Zhang, M. van Vugt, J.P. Borst, J.R. Anderson, Mapping working memory retrieval in space and in time: a combined electroencephalography and electrocorticography approach, Neuroimage 174 (2018) 472–484, https://doi.org/10.1016/j.neuroimage.2018.03.039.
[137] J.G. Samuelsson, P. Sundaram, S. Khan, M.I. Sereno, M.S. Hämäläinen, Detectability of cerebellar activity with magnetoencephalography and electroencephalography, Hum. Brain Mapp. 41 (2020) 2357–2372, https://doi.org/10.1002/hbm.24951.
[138] A.R. Levin, A.S. Méndez Leal, L.J. Gabard-Durnam, H.M. O'Leary, BEAPP: the batch electroencephalography automated processing platform, Front. Neurosci. 12 (2018) 513, https://doi.org/10.3389/fnins.2018.00513.
[139] C.R. Pernet, S. Appelhoff, K.J. Gorgolewski, G. Flandin, C. Phillips, A. Delorme, R. Oostenveld, EEG-BIDS, an extension to the brain imaging data structure for electroencephalography, Sci. Data 6 (2019) 103, https://doi.org/10.1038/s41597-019-0104-8.
[140] M. Hassib, S. Schneegass, P. Eiglsperger, N. Henze, A. Schmidt, F. Alt, EngageMeter: a system for implicit audience engagement sensing using electroencephalography, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, Association for Computing Machinery, 2017, pp. 5114–5119, https://doi.org/10.1145/3025453.3025669.
[141] M. Hassib, M. Pfeiffer, S. Schneegass, M. Rohs, F. Alt, Emotion actuator: embodied emotional feedback through electroencephalography and electrical muscle stimulation, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, Association for Computing Machinery, 2017, pp. 6133–6146, https://doi.org/10.1145/3025453.3025953.
[142] M. Hassib, M. Khamis, S. Friedl, S. Schneegass, F. Alt, Brainatwork: logging cognitive engagement and tasks in the workplace using electroencephalography, in: Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia, New York, NY, USA, Association for Computing Machinery, 2017, pp. 305–310, https://doi.org/10.1145/3152832.3152865.
[143] L.J. Gabard-Durnam, A.S. Mendez Leal, C.L. Wilkinson, A.R. Levin, The Harvard automated processing pipeline for electroencephalography (HAPPE): standardized processing software for developmental and high-artifact data, Front. Neurosci. 12 (2018) 97, https://doi.org/10.3389/fnins.2018.00097.
[144] C.T. Lin, Y.H. Yu, J.T. King, C.H. Liu, L. De Liao, Augmented wire-embedded silicon-based dry-contact sensors for electroencephalography signal measurements, IEEE Sens. J. 20 (2020), https://doi.org/10.1109/JSEN.2019.2959619.
[145] S. Wan, Z. Gu, Q. Ni, Cognitive computing and wireless communications on the edge for healthcare service robots, Comput. Commun. 149 (2020) 99–106, https://doi.org/10.1016/j.comcom.2019.10.012.
[146] S. Chakraborty, Wireless body area sensor network in healthcare applications, in: SoutheastCon, 2018, 2018, pp. 1–2, https://doi.org/10.1109/SECON.2018.8479124.
[147] M.A. Matin, M.M. Islam, Overview of wireless sensor network. Wireless Sensor Networks - Technology and Protocols, InTech, 2012, https://doi.org/10.5772/49376.
[148] K. Açıcı, Ç.B. Erdaş, T. Aşuroğlu, M.K. Toprak, H. Erdem, H. Oğul, A random forest method to detect Parkinson's disease via gait analysis. Communications in Computer and Information Science, Springer Verlag, 2017, pp. 609–619, https://doi.org/10.1007/978-3-319-65172-9_51.
[149] S. Chakraborty, A. Jamthe, S.K. Ghosh, D.P. Agrawal, From theory to application: wireless monitoring of patients suffering from neurodegenerative diseases, in: Midwest Symposium on Circuits and Systems, 2013, pp. 944–947, https://doi.org/10.1109/MWSCAS.2013.6674806.
[150] S. Bringas, S. Salomón, R. Duque, J.L. Montaña, C. Lage, A convolutional neural network-based method for human movement patterns classification in Alzheimer's disease, Proceedings 31 (2019), https://doi.org/10.3390/proceedings2019031072.
[151] S. Moon, H.-J. Song, V.D. Sharma, K.E. Lyons, R. Pahwa, A.E. Akinwuntan, H. Devos, Classification of Parkinson's disease and essential tremor based on balance and gait characteristics from wearable motion sensors via machine learning techniques: a data-driven approach, J. Neuroeng. Rehabil. 17 (2020) 125, https://doi.org/10.1186/s12984-020-00756-5.
[152] M. Malcangi, AI-based methods and technologies to develop wearable devices for prosthetics and predictions of degenerative diseases. Methods in Molecular Biology, Humana Press Inc., 2021, pp. 337–354, https://doi.org/10.1007/978-1-0716-0826-5_17.
[153] E. Rastegari, S. Azizian, H. Ali, Machine learning and similarity network approaches to support automatic classification of Parkinson's diseases using accelerometer-based gait analysis, in: Proceedings of the 52nd Hawaii International Conference on System Sciences, Hawaii International Conference on System Sciences, 2019, https://doi.org/10.24251/HICSS.2019.511.
[154] K. Gupta, A. Khajuria, N. Chatterjee, P. Joshi, D. Joshi, Rule based classification of neurodegenerative diseases using data driven gait features, Health Technol. (Berl) 9 (2019) 547–560, https://doi.org/10.1007/s12553-018-0274-y.
[155] N. Nahar, M.S. Hossain, K. Andersson, A machine learning based fall detection for elderly people with neurodegenerative disorders. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Science and Business Media Deutschland GmbH, 2020, pp. 194–203, https://doi.org/10.1007/978-3-030-59277-6_18.
[156] T. Matkovič, V. Pejović, Wi-Mind: wireless mental effort inference, in: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, New York, NY, USA, Association for Computing Machinery, 2018, pp. 1241–1249, https://doi.org/10.1145/3267305.3274114.
[157] S. Gheisari, M.R. Meybodi, A new reasoning and learning model for cognitive wireless sensor networks based on Bayesian networks and learning automata cooperation, Comput. Netw. 124 (2017) 11–26, https://doi.org/10.1016/j.comnet.2017.05.031.
[158] A. Depari, D. Fernandes Carvalho, P. Bellagente, P. Ferrari, E. Sisinni, A. Flammini, A. Padovani, An IoT based architecture for enhancing the effectiveness of prototype medical instruments applied to neurodegenerative disease diagnosis, Sensors 19 (2019), https://doi.org/10.3390/s19071564.
[159] I. García-Magariño, J. Varela-Aldas, G. Palacios-Navarro, J. Lloret, Fog computing for assisting and tracking elder patients with neurodegenerative diseases, Peer Peer Netw. Appl. 12 (2019) 1225–1235, https://doi.org/10.1007/s12083-019-00732-4.
[160] I.M. Saied, S. Chandran, T. Arslan, Integrated flexible hybrid silicone-textile dual-resonant sensors and switching circuit for wearable neurodegeneration monitoring systems, IEEE Trans. Biomed. Circuits Syst. 13 (2019) 1304–1312, https://doi.org/10.1109/TBCAS.2019.2951500.
[161] A. Dvorani, C. Wiesener, M. Valtin, H. Voigt, A. Kühn, N. Wenger, T. Schauer, Mobil4Park: development of a sensor-stimulator network for the therapy of freezing of gait in Parkinson patients, Curr. Dir. Biomed. Eng. 6 (2020), https://doi.org/10.1515/cdbme-2020-2013.
[162] D. Bastos, J. Ribeiro, F. Silva, M. Rodrigues, A.G. Silva, A. Queirós, A. Fernández-Caballero, N.P. Rocha, A. Pereira, SmartWalk BAN: using body area networks to encourage older adults to perform physical activity, Electronics (Basel) 10 (2021), https://doi.org/10.3390/electronics10010056.
[163] K. Kalovrektis, A. Xenakis, I. Korinthios, A. Kakarountas, A. Gotsinas, G. Stamoulis, An IoT -sensor based hand trembling robotic system for studying Parkinson's symptoms, 9 (2020) 12–19. https://doi.org/10.5923/j.ijit.20200901.02.
[164] W. Tao, T. Liu, R. Zheng, H. Feng, Gait analysis using wearable sensors, Sensors (Basel) 12 (2012) 2255–2283, https://doi.org/10.3390/s120202255.
[165] G. Gugliandolo, G. Campobello, P.P. Capra, S. Marino, A. Bramanti, G. Di Lorenzo, N. Donato, a movement-tremors recorder for patients of neurodegenerative diseases, IEEE Trans. Instrum. Meas. 68 (2019) 1451–1457, https://doi.org/10.1109/TIM.2019.2900141.
[166] M.J. Kim, E. Naydanova, B.Y. Hwang, K.A. Mills, W.S. Anderson, Y. Salimpour, Quantification of Parkinson's disease motor symptoms: a wireless motion sensing approach, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 3658–3661, https://doi.org/10.1109/EMBC44109.2020.9175616.
[167] R. LeMoyne, T. Mastroianni, D. Whiting, N. Tomycz, Wearable and wireless systems with internet connectivity for quantification of Parkinson's disease and essential tremor characteristics. Smart Sensors, Measurement and Instrumentation, Springer International Publishing, 2019, pp. 79–97, https://doi.org/10.1007/978-981-13-5808-1_7.

[168] T. Zia, A. Zomaya, Security issues in wireless sensor networks, in: 2006 International Conference on Systems and Networks Communications (ICSNC'06), 2006, p. 40, https://doi.org/10.1109/ICSNC.2006.66.
[169] J. Portilla, A. Otero, V. Rosello, J. Valverde, Y.E. Krasteva, E. de la Torre, T. Riesgo, 13.13 - Wireless Sensor Networks: From Real World to System Integration – Alternative Hardware Approaches, in: S. Hashmi, G.F. Batalha, C.J. Van Tyne, B.B.T.-C.M.P. Yilbas (Eds.), Elsevier, Oxford, 2014: pp. 353–373. 10.1016/B978-0-08-096532-1.01313-3.
[170] S. Radhika, P. Rangarajan, On improving the lifespan of wireless sensor networks with fuzzy based clustering and machine learning based data reduction, Appl. Soft Comput. 83 (2019) 105610, https://doi.org/10.1016/j.asoc.2019.105610.
[171] B.P. Selgrade, D. Meyer, J.J. Sosnoff, J.R. Franz, Can optical flow perturbations detect walking balance impairment in people with multiple sclerosis? PLoS ONE 15 (2020) 1–16, https://doi.org/10.1371/journal.pone.0230202.
[172] M. Buegler, R.L. Harms, M. Balasa, I.B. Meier, T. Exarchos, L. Rai, R. Boyle, A. Tort, M. Kozori, E. Lazarou, M. Rampini, C. Cavaliere, P. Vlamos, M. Tsolaki, C. Babiloni, A. Soricelli, G. Frisoni, R. Sanchez-Valle, R. Whelan, E. Merlo-Pich, I. Tarnanas, Digital biomarker-based individualized prognosis for people at risk of dementia, Alzheimer's and dementia, Diagn. Assess. Dis. Monit. 12 (2020) 1–13, https://doi.org/10.1002/dad2.12073.
[173] D. Howett, A. Castegnaro, K. Krzywicka, J. Hagman, D. Marchment, R. Henson, M. Rio, J.A. King, N. Burgess, D. Chan, Differentiation of mild cognitive impairment using an entorhinal cortex-based test of virtual reality navigation, Brain 142 (2019) 1751–1766, https://doi.org/10.1093/brain/awz116.
[174] G. Albani, S. Raspelli, L. Carelli, L. Priano, R. Pignatti, F. Morganti, A. Gaggioli, P.L. Weiss, R. Kizony, N. Katz, A. Mauro, G. Riva, Sleep dysfunctions influence decision making in undemented Parkinson's disease patients: a study in a virtual supermarket, Stud. Health Technol. Inform. 163 (2011) 8–10.
[175] J. Orlosky, Y. Itoh, M. Ranchet, K. Kiyokawa, J. Morgan, H. Devos, Emulation of physician tasks in eye-tracked virtual reality for remote diagnosis of neurodegenerative disease, IEEE Trans. Vis. Comput. Graph 23 (2017) 1302–1311, https://doi.org/10.1109/TVCG.2017.2657018.
[176] C. Versek, A. Rissmiller, A. Tran, M. Taya, K. Chowdhury, P. Bex, S. Sridhar, Portable system for neuro-optical diagnostics using virtual reality display, Mil Med. 184 (2019) 584–592, https://doi.org/10.1093/milmed/usy286.
[177] H. Ahmadi, G. Arji, L. Shahmoradi, R. Safdari, M. Nilashi, M. Alizadeh, The Application of Internet of Things in Healthcare: a Systematic Literature Review and Classification, Springer, Berlin Heidelberg, 2019, https://doi.org/10.1007/s10209-018-0618-4.
[178] S.P. Dash, The impact of IoT in healthcare: global technological change & the roadmap to a networked architecture in India, J. Indian Inst. Sci. 100 (2020) 773–785, https://doi.org/10.1007/s41745-020-00208-y.
[179] S. Sharma, R.K. Dudeja, G.S. Aujla, R.S. Bali, N. Kumar, DeTrAs: deep learning-based healthcare framework for IoT-based assistance of Alzheimer patients, Neural Comput. Appl. 2 (2020), https://doi.org/10.1007/s00521-020-05327-2.
[180] T.G. Stavropoulos, A. Papastergiou, L. Mpaltadoros, S. Nikolopoulos, I. Kompatsiaris, Iot wearable sensors and devices in elderly care: a literature review, Sensors (Switzerland) 20 (2020), https://doi.org/10.3390/s20102826.
[181] A. Sciarrone, I. Bisio, C. Garibotto, F. Lavagetto, G.H. Staude, A. Knopp, Leveraging IoT wearable technology towards early diagnosis of neurological diseases, IEEE J. Sel. Areas Commun. 39 (2021) 582–592, https://doi.org/10.1109/JSAC.2020.3021573.
[182] Z. Lv, Virtual reality in the context of internet of things, Neural Comput. Appl. 32 (2020) 9593–9602, https://doi.org/10.1007/s00521-019-04472-7.
[183] Y. Park, S. Yun, K.-H. Kim, When IoT met augmented reality, (2019) 117–129. https://doi.org/10.1145/3307334.3326079.
[184] F. Ghorbani, M. Kia, M. Delrobaei, Q. Rahman, Evaluating the possibility of integrating augmented reality and internet of things technologies to help patients with Alzheimer's disease, in: 2019 26th National and 4th International Iranian Conference on Biomedical Engineering, ICBME 2019, 2019, pp. 139–144, https://doi.org/10.1109/ICBME49163.2019.9030404.
[185] Z. Rashid, J. Melià-Seguí, R. Pous, E. Peig, Using augmented reality and internet of things to improve accessibility of people with motor disabilities in the context of smart cities, Future Gener. Comput. Syst. 76 (2017) 248–261, https://doi.org/10.1016/j.future.2016.11.030.
[186] M. Ben-Moussa, M. Rubo, C. Debracque, W.G. Lange, DJInnI: a novel technology supported exposure therapy paradigm for SAD combining virtual reality and augmented reality, Front. Psychiatry 8 (2017), https://doi.org/10.3389/fpsyt.2017.00026.
[187] G. Higginbotham, Virtual connections: improving global neurosurgery through immersive technologies, Front. Surg. 8 (2021) 629963, https://doi.org/10.3389/fsurg.2021.629963.
[188] D.W. Roberts, J.W. Strohbehn, J.F. Hatch, W. Murray, H. Kettenberger, A frameless stereotaxic integration of computerized tomographic imaging and the operating microscope, J. Neurosurg. 65 (1986) 545–549, https://doi.org/10.3171/jns.1986.65.4.0545.
[189] P.L. Gleason, R. Kikinis, D. Altobelli, W. Wells, E. Alexander, P.M.L. Black, F. Jolesz, Video registration virtual reality for nonlinkage stereotactic surgery, Stereotact. Funct. Neurosurg. 63 (1994) 139–143, https://doi.org/10.1159/000100305.
[190] P.J. Edwards, D.J. Hawkes, D.L.G. Hill, D. Jewell, R. Spink, A. Strong, M. Gleeson, Augmentation of reality using an operating microscope for otolaryngology and neurosurgical guidance, J. Image Guid. Surg. 1 (1995) 172–178, https://doi.org/10.1002/(sici)1522-712x(1995)1:3<172::aid-igs7>3.0.co;2-7.
[191] Y. Masutani, T. Dohi, F. Yamane, H. Iseki, K. Takakura, Augmented reality visualization system for intravascular neurosurgery, Comput. Aided Surg. 3 (1998) 239–247, https://doi.org/10.1002/(sici)1097-0150(1998)3:5<239::aid-igs3>3.0.co;2-b.
[192] T. Kawamata, H. Iseki, T. Shibasaki, T. Hori, C.B. Heilman, G.H. Barnett, M.L. Levy, Endoscopic augmented reality navigation system for endonasal transsphenoidal surgery to treat pituitary tumors: technical note, Neurosurgery 50 (2002) 1393–1397, https://doi.org/10.1097/00006123-200206000-00038.
[193] D. Guha, N.M. Alotaibi, N. Nguyen, S. Gupta, C. McFaul, V.X.D. Yang, Augmented reality in neurosurgery: a review of current concepts and emerging applications, Can. J. Neurol. Sci. 44 (2017) 235–245, https://doi.org/10.1017/cjn.2016.443.
[194] R. Tagaytayan, A. Kelemen, C. Sik-Lanyi, Augmented reality in neurosurgery, Arch. Med. Sci. 14 (2018) 572–578, https://doi.org/10.5114/aoms.2016.58690.
[195] Z. Zhao, P. Niu, X. Ji, R.M. Sweet, State of simulation in healthcare education: an initial survey in Beijing, J. Soc. Laparoendosc. Surg. 21 (2017), https://doi.org/10.4293/JSLS.2016.00090.
[196] S.O. Mahboob, M. Eljamel, Intraoperative image-guided surgery in neuro-oncology with specific focus on high-grade gliomas, Future Oncol. 13 (2017) 2349–2361, https://doi.org/10.2217/fon-2017-0195.
[197] M. Marigil, M. Bernstein, Outpatient neurosurgery in neuro-oncology, Neurosurg. Focus 44 (2018) E19, https://doi.org/10.3171/2018.3.FOCUS1831.
[198] W. Deng, F. Li, M. Wang, Z. Song, Easy-to-use augmented reality neuronavigation using a wireless tablet PC, Stereotact. Funct. Neurosurg. 92 (2014) 17–24, https://doi.org/10.1159/000354816.
[199] W.L. Nowinski, Q.L. Yang, T.T. Yeo, Computer-aided stereotactic functional neurosurgery enhanced by the use of the multiple brain atlas database, IEEE Trans. Med. Imaging 19 (2000) 62–69, https://doi.org/10.1109/42.832961.
[200] B. Fiani, F. De Stefano, A. Kondilis, C. Covarrubias, L. Reier, K. Sarhadi, Virtual reality in neurosurgery: "Can you see it?"–A review of the current applications and future potential, World Neurosurg. 141 (2020) 291–298, https://doi.org/10.1016/j.wneu.2020.06.066.
[201] F. Cutolo, A. Meola, M. Carbone, S. Sinceri, F. Cagnazzo, E. Denaro, N. Esposito, M. Ferrari, V. Ferrari, A new head-mounted display-based augmented reality system in neurosurgical oncology: a study on phantom, Comput. Assist. Surg. 22 (2017) 39–53, https://doi.org/10.1080/24699322.2017.1358400.
[202] D. Inoue, B. Cho, M. Mori, Y. Kikkawa, T. Amano, A. Nakamizo, K. Yoshimoto, M. Mizoguchi, M. Tomikawa, J. Hong, M. Hashizume, T. Sasaki, Preliminary study on the clinical application of augmented reality neuronavigation, J. Neurol. Surg. A Cent. Eur. Neurosurg. 74 (2013) 71–76, https://doi.org/10.1055/s-0032-1333415.
[203] M. Mahvash, L.B. Tabrizi, A novel augmented reality system of image projection for image-guided neurosurgery, Acta Neurochir. (Wien) 155 (2013) 943–947, https://doi.org/10.1007/s00701-013-1668-2.
[204] S. Drouin, A. Kochanowska, M. Kersten-Oertel, I.J. Gerard, R. Zelmann, D. De Nigris, S. Bériault, T. Arbel, D. Sirhan, A.F. Sadikot, J.A. Hall, D.S. Sinclair, K. Petrecca, R.F. DelMaestro, D.L. Collins, IBIS: an OR ready open-source platform for image-guided neurosurgery, Int. J. Comput. Assist. Radiol. Surg. 12 (2017) 363–378, https://doi.org/10.1007/s11548-016-1478-0.
[205] P.L. Gildenberg, J. Labuz, Use of a volumetric target for image-guided surgery, Neurosurgery 59 (2006) 651–658, https://doi.org/10.1227/01.NEU.0000227474.21048.F1.
[206] I. Cabrilo, K. Schaller, P. Bijlenga, Augmented reality-assisted bypass surgery: embracing minimal invasiveness, World Neurosurg. 83 (2015) 596–602, https://doi.org/10.1016/j.wneu.2014.12.020.

[207] M.C. Brodsky, Neuro-ophthalmologic manifestations of systemic and intracranial disease. Pediatric Neuro-Ophthalmology, Springer, New York, 2016, pp. 649–776, https://doi.org/10.1007/978-1-4939-3384-6_11.
[208] M. Kersten-Oertel, S.S.J. Chen, S. Drouin, D.S. Sinclair, D.L. Collins, Augmented reality visualization for guidance in neurovascular surgery. Studies in Health Technology and Informatics, IOS Press, 2012, pp. 225–229, https://doi.org/10.3233/978-1-61499-022-2-225.
[209] C.S. Ong, G. Deib, P. Yesantharao, Y. Qiao, J. Pakpoor, N. Hibino, F. Hui, J.R. Garcia, Virtual reality in neurointervention, J. Vasc. Interv. Neurol. 10 (2018) 17–22.
[210] A. Alaraj, C.J. Luciano, D.P. Bailey, A. Elsenousi, B.Z. Roitberg, A. Bernardo, P.P. Banerjee, F.T. Charbel, Virtual reality cerebral aneurysm clipping simulation with real-time haptic feedback, Oper. Neurosur. 11 (2015) 52–58, https://doi.org/10.1227/NEU.0000000000000583.
[211] M.B. Shenai, M. Dillavou, C. Shum, D. Ross, R.S. Tubbs, A. Shih, B.L. Guthrie, Virtual interactive presence and augmented reality (VIPAR) for remote surgical assistance, Neurosurgery 68 (2011), https://doi.org/10.1227/NEU.0b013e3182077efd.
[212] M. Kersten-Oertel, I. Gerard, S. Drouin, K. Mok, D. Sirhan, D.S. Sinclair, D.L. Collins, Augmented reality in neurovascular surgery: feasibility and first uses in the operating room, Int. J. Comput. Assist. Radiol. Surg. 10 (2015) 1823–1836, https://doi.org/10.1007/s11548-015-1163-8.
[213] Y. Abe, S. Sato, K. Kato, T. Hyakumachi, Y. Yanagibashi, M. Ito, K. Abumi, A novel 3D guidance system using augmented reality for percutaneous vertebroplasty, J. Neurosurg. Spine 19 (2013) 492–501, https://doi.org/10.3171/2013.7.SPINE12917.
[214] M. Pfandler, M. Lazarovici, P. Stefan, P. Wucherer, M. Weigl, Virtual reality-based simulators for spine surgery: a systematic review, Spine J. 17 (2017) 1352–1363, https://doi.org/10.1016/j.spinee.2017.05.016.
[215] B.J. Shin, A.R. James, I.U. Njoku, R. Härtl, Pedicle screw navigation: a systematic review and meta-analysis of perforation risk for computer-navigated versus freehand insertion - a review, J. Neurosurg. Spine 17 (2012) 113–122, https://doi.org/10.3171/2012.5.SPINE11399.
[216] N. Navab, S.M. Heining, J. Traub, Camera augmented mobile C-arm (CAMC): calibration, accuracy study, and clinical applications, IEEE Trans. Med. Imaging 29 (2010) 1412–1423, https://doi.org/10.1109/TMI.2009.2021947.
[217] B.J. Park, S.J. Hunt, G.J. Nadolski, T.P. Gade, Augmented reality improves procedural efficiency and reduces radiation dose for CT-guided lesion targeting: a phantom study using HoloLens 2, Sci. Rep. 10 (2020) 1–8, https://doi.org/10.1038/s41598-020-75676-4.
[218] F. Heinrich, L. Schwenderling, M. Becker, M. Skalej, C. Hansen, Holoinjection: augmented reality support for CT-guided spinal needle injections, Healthc. Technol. Lett. (2019) 165–171, https://doi.org/10.1049/htl.2019.0062.
[219] J.R. Wu, M.L. Wang, K.C. Liu, M.H. Hu, P.Y. Lee, Real-time advanced spinal surgery via visible patient model and augmented reality system, Comput. Methods Programs Biomed. 113 (2014) 869–881, https://doi.org/10.1016/j.cmpb.2013.12.021.
[220] J. Gasco, A. Patel, J. Ortega-Barnett, D. Branch, S. Desai, Y.F. Kuo, C. Luciano, S. Rizzi, P. Kania, M. Matuyauskas, P. Banerjee, B.Z. Roitberg, Virtual reality spine surgery simulation: an empirical study of its usefulness, Neurol. Res. 36 (2014) 968–973, https://doi.org/10.1179/1743132814Y.0000000388.
[221] E. Moult, T. Ungi, M. Welch, J. Lu, R.C. McGraw, G. Fichtinger, Ultrasound-guided facet joint injection training using Perk Tutor, Int. J. Comput. Assist. Radiol. Surg. 8 (2013) 831–836, https://doi.org/10.1007/s11548-012-0811-5.
[222] A.T.F. Da Silva, J.A.S. Carneiro, L.C. De Oliveira, A.D. Amaral, M.M. Da Silva, R.G. Andrade, V.A. Pereira, Serious game for chronic subdural hematoma neurosurgy using virtual reality, in: Proceedings - 2020 22nd Symposium on Virtual and Augmented Reality, SVR 2020, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 381–385, https://doi.org/10.1109/SVR51698.2020.00063.
[223] A.T. Stadie, R.A. Kockro, R. Reisch, A. Tropine, S. Boor, P. Stoeter, A. Perneczky, Virtual reality system for planning minimally invasive neurosurgery: technical note, J. Neurosurg. 108 (2008) 382–394, https://doi.org/10.3171/JNS/2008/108/2/0382.
[224] P. Ferroli, G. Tringali, F. Acerbi, M. Schiariti, M. Broggi, D. Aquino, G. Broggi, Advanced 3-dimensional planning in neurosurgery, Neurosurgery 72 (2013) A54–A62, https://doi.org/10.1227/NEU.0b013e3182748ee8.
[225] R.A. Kockro, T. Killeen, A. Ayyad, M. Glaser, A. Stadie, R. Reisch, A. Giese, E. Schwandt, Aneurysm surgery with preoperative three-dimensional planning in a virtual reality environment: technique and outcome analysis, World Neurosurg. 96 (2016) 489–499, https://doi.org/10.1016/j.wneu.2016.08.124.
[226] A. Alaraj (Ed.), Comprehensive Healthcare Simulation: Neurosurgery, Springer International Publishing, Cham, 2018, https://doi.org/10.1007/978-3-319-75583-0.
[227] B. Dalgarno, M.J.W. Lee, What are the learning affordances of 3-D virtual environments? Br. J. Educ. Technol. 41 (2010) 10–32, https://doi.org/10.1111/j.1467-8535.2009.01038.x.
[228] P.E. Pelargos, D.T. Nagasawa, C. Lagman, S. Tenn, J.V. Demos, S.J. Lee, T.T. Bui, N.E. Barnette, N.S. Bhatt, N. Ung, A. Bari, N.A. Martin, I. Yang, Utilizing virtual and augmented reality for educational and clinical enhancements in neurosurgery, J. Clin. Neurosci. 35 (2017) 1–4, https://doi.org/10.1016/j.jocn.2016.09.002.
[229] N. Choudhury, N. Gélinas-Phaneuf, S. Delorme, R. Del Maestro, Fundamentals of neurosurgery: virtual reality tasks for training and evaluation of technical skills, World Neurosurg. 80 (2013) e9–e19, https://doi.org/10.1016/j.wneu.2012.08.022.
[230] M. Nakamoto, O. Ukimura, I.S. Gill, A. Mahadevan, T. Miki, M. Hashizume, Y. Sato, Realtime organ tracking for endoscopic augmented reality visualization using miniature wireless magnetic tracker. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer, Berlin, Heidelberg, 2008, pp. 359–366, https://doi.org/10.1007/978-3-540-79982-5_39.
[231] D. Teber, S. Guven, T. Simpfendörfer, M. Baumhauer, E.O. Güven, F. Yencilek, A.S. Gözen, J. Rassweiler, Augmented reality: a new tool to improve surgical accuracy during laparoscopic partial nephrectomy? Preliminary in vitro and in vivo results, Eur. Urol. 56 (2009) 332–338, https://doi.org/10.1016/j.eururo.2009.05.017.
[232] I. Cabrilo, A. Sarrafzadeh, P. Bijlenga, B.N. Landis, K. Schaller, Augmented reality-assisted skull base surgery, Neurochirurgie 60 (2014) 304–306, https://doi.org/10.1016/j.neuchi.2014.07.001.
[233] X. Zang, J. Yang, D. Weng, Y. Wang, Y. Liu, Augmented reality based surgery navigation system, in: T. Yoshizawa, P. Wei, J. Zheng (Eds.), 2009 International Conference on Optical Instruments and Technology: Optoelectronic Imaging and Process Technology, SPIE, 2009, p. 75132T, https://doi.org/10.1117/12.839658.
[234] R.U. Thoranaghatte, J.G. Giraldez, G. Zheng, Landmark based augmented reality endoscope system for sinus and skull-base surgeries, in: Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'08 - "Personalized Healthcare through Technology, IEEE Computer Society, 2008, pp. 74–77, https://doi.org/10.1109/iembs.2008.4649094.
[235] A.R. M, V.M. H, M.S. JM, T.S. R, J.L. MC, M.A. C, G.G. JA, M. Alcaniz Raya, H. Varvaro Marcinek, J.M. Martinez Saez, R. Torres Sanchez, M.C. Juan Lizandra, C. Monserrat Aranda, J.A. Gil Gomez, Mixed reality for neurosurgery: a novel prototype, Stud. Health Technol. Inform. 94 (2003) 11–15.
[236] L.B. Tabrizi, M. Mahvash, Augmented reality-guided neurosurgery: accuracy and intraoperative application of an image projection technique, J. Neurosurg. 123 (2015) 206–211, https://doi.org/10.3171/2014.9.JNS141001.
[237] A. Meola, F. Cutolo, M. Carbone, F. Cagnazzo, M. Ferrari, V. Ferrari, Augmented reality in neurosurgery: a systematic review, Neurosurg. Rev. 40 (2017) 537–548, https://doi.org/10.1007/s10143-016-0732-9.
[238] M. Aschke, C.R. Wirtz, J. Raczkowsky, H. Wörn, S. Kunze, Augmented reality in operating microscopes for neurosurgical interventions, in: International IEEE/EMBS Conference on Neural Engineering, NER, IEEE Computer Society, 2003, pp. 652–655, https://doi.org/10.1109/CNE.2003.1196913.
[239] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, B. MacIntyre, Recent advances in augmented reality, IEEE Comput. Graph Appl. 21 (2001) 34–47, https://doi.org/10.1109/38.963459.
[240] J. Cho, S. Rahimpour, A. Cutler, C.R. Goodwin, S.P. Lad, P. Codd, Enhancing reality: a systematic review of augmented reality in neuronavigation and education, World Neurosurg. 139 (2020) 186–195, https://doi.org/10.1016/j.wneu.2020.04.043.
[241] F. Zhang, L. Chen, W. Miao, L. Sun, Research on accuracy of augmented reality surgical navigation system based on multi-view virtual and real registration technology, IEEE Access 8 (2020) 122511–122528, https://doi.org/10.1109/ACCESS.2020.3007178.
[242] V. Pennacchietti, K. Stoelzel, A. Tietze, E. Lankes, A. Schaumann, F.C. Uecker, U.W. Thomale, First experience with augmented reality neuronavigation in endoscopic assisted midline skull base pathologies in children, Child's Nerv. Syst. (2021), https://doi.org/10.1007/s00381-021-05049-3.

[243] J. Haemmerli, A. Davidovic, T.R. Meling, L. Chavaz, K. Schaller, P. Bijlenga, Evaluation of the precision of operative augmented reality compared to standard neuronavigation using a 3D-printed skull, Neurosurg. Focus FOC. 50 (2021) E17, https://doi.org/10.3171/2020.10.FOCUS20789.
[244] I.J. Gerard, M. Kersten-Oertel, S. Drouin, J.A. Hall, K. Petrecca, D. De Nigris, T. Arbel, D.L. Collins, Improving patient specific neurosurgical models with intraoperative ultrasound and augmented reality visualizations in a neuronavigation environment. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 2016, pp. 28–35, https://doi.org/10.1007/978-3-319-31808-0_4.
[245] K. Maruyama, E. Watanabe, K. Saito, K. Sato, H. Yoshida, S. Kume, A. Noguchi, Y. Shiokawa, M. Nagane, P04.17 Novel augmented reality-based neuronavigation using smart glasses and clinical application to brain tumor surgery, Neuro Oncol. 19 (2017) iii43–iii44, https://doi.org/10.1093/neuonc/nox036.157.
[246] K. Maruyama, E. Watanabe, T. Kin, K. Saito, A. Kumakiri, A. Noguchi, M. Nagane, Y. Shiokawa, Smart glasses for neurosurgical navigation by augmented reality, Oper. Neurosurg. 15 (2018) 551–556, https://doi.org/10.1093/ons/opx279.
[247] T. Fick, J.A.M. van Doormaal, E.W. Hoving, L. Regli, T.P.C. van Doormaal, Holographic patient tracking after bed movement for augmented reality neuronavigation using a head-mounted display, Acta Neurochir. (Wien) 163 (2021) 879–884, https://doi.org/10.1007/s00701-021-04707-4.
[248] C. Leuze, G. Yang, B. Hargreaves, B. Daniel, J.A. McNab, Mixed-reality guidance for brain stimulation treatment of depression, in: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), 2018, pp. 377–380, https://doi.org/10.1109/ISMAR-Adjunct.2018.00109.
[249] S. Sathyanarayana, C. Leuze, B. Hargreaves, B. Daniel, G. Wetzstein, A. Etkin, M.T. Bhati, J.A. McNab, Comparison of head pose tracking methods for mixed-reality neuronavigation for transcranial magnetic stimulation, in: Proc. SPIE, 2020, https://doi.org/10.1117/12.2547917.
[250] A. Gronningsaeter, A. Kleven, S. Ommedal, T.E. Aarseth, T. Lie, F. Lindseth, T. Langø, G. Unsgård, SonoWand, an ultrasound-based neuronavigation system, Neurosurgery 47 (2000) 1373–1380, https://doi.org/10.1097/00006123-200012000-00021.
[251] M.H.T. Reinges, H.-H. Nguyen, T. Krings, B.-O. Hütter, V. Rohde, J.M. Gilsbach, Course of brain shift during microsurgical resection of supratentorial cerebral lesions: limits of conventional neuronavigation, Acta Neurochir. (Wien) 146 (2004) 369–377, https://doi.org/10.1007/s00701-003-0204-1.
[252] A.T. Stadie, R.A. Kockro, L. Serra, G. Fischer, E. Schwandt, P. Grunert, R. Reisch, Neurosurgical craniotomy localization using a virtual reality planning system versus intraoperative image–guided navigation, Int. J. Comput. Assist. Radiol. Surg. 6 (2011) 565–572, https://doi.org/10.1007/s11548-010-0529-1.
[253] B.F.G. Katz, S. Kammoun, G. Parseihian, O. Gutierrez, A. Brilhault, M. Auvray, P. Truillet, M. Denis, S. Thorpe, C. Jouffrais, NAVIG: augmented reality guidance system for the visually impaired, Virtual Real. 16 (2012) 253–269, https://doi.org/10.1007/s10055-012-0213-6.
[254] Y.-J. Chang, Y.-S. Kang, Y.-S. Chang, H.-H. Liu, ARCoach 2.0: optimizing a vocational prompting system based on augmented reality for people with cognitive impairments, in: Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility, New York, NY, USA, Association for Computing Machinery, 2015, pp. 313–314, https://doi.org/10.1145/2700648.2811354.
[255] M. Rossi, G. D'Avenio, S. Morelli, M. Grigioni, CogAR: an augmented reality app to improve quality of life of the people with cognitive impairment, in: 2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON), 2020, pp. 339–343, https://doi.org/10.1109/MELECON48756.2020.9140554.
[256] C.R. Oliveira, B.J.P. Lopes Filho, M.A. Sugarman, C.S. Esteves, M.M.B.M.P. Lima, C. Moret-Tatay, T.Q. Irigaray, I.I.L. Argimon, Development and feasibility of a virtual reality task for the cognitive assessment of older adults: the ECO-VR, Span. J. Psychol. 19 (2016) E95, https://doi.org/10.1017/sjp.2016.96.
[257] D. Wolf, D. Besserer, K. Sejunaite, A. Schuler, M. Riepe, E. Rukzio, CARe: an augmented reality support system for geriatric inpatients with mild cognitive impairment, in: Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia, New York, NY, USA, Association for Computing Machinery, 2019, https://doi.org/10.1145/3365610.3365612.
[258] C.-Y. Lin, Y.-M. Chang, Interactive augmented reality using Scratch 2.0 to improve physical activities for children with developmental disabilities, Res. Dev. Disabil. 37 (2015) 1–8, https://doi.org/10.1016/j.ridd.2014.10.016.
[259] H.E. Boz, H. Limoncu, S. Zygouris, M. Tsolaki, D. Giakoumis, K. Votis, D. Tzovaras, V. Öztürk, G.G. Yener, A new tool to assess amnestic mild cognitive impairment in Turkish older adults: virtual supermarket (VSM), Aging, Neuropsychol. Cogn. 27 (2020) 639–653, https://doi.org/10.1080/13825585.2019.1663146.
[260] J. Park, J. Yim, A new approach to improve cognition, muscle strength, and postural balance in community-dwelling elderly with a 3-D virtual reality kayak program, Tohoku J. Exp. Med. 238 (2016) 1–8, https://doi.org/10.1620/tjem.238.1.
[261] A. Mirelman, L. Rochester, M. Reelick, F. Nieuwhof, E. Pelosin, G. Abbruzzese, K. Dockx, A. Nieuwboer, J.M. Hausdorff, V-TIME: a treadmill training program augmented by virtual reality to decrease fall risk in older adults: study design of a randomized controlled trial, BMC Neurol. 13 (2013) 15, https://doi.org/10.1186/1471-2377-13-15.
[262] P. Cipresso, G. Albani, S. Serino, E. Pedroli, F. Pallavicini, A. Mauro, G. Riva, Virtual multiple errands test (VMET): a virtual reality-based tool to detect early executive functions deficit in Parkinson's disease, Front. Behav. Neurosci. 8 (2014) 405, https://doi.org/10.3389/fnbeh.2014.00405.
[263] L.P. Carvalho, K.K.V Mate, E. Cinar, A. Abou-Sharkh, A.-L. Lafontaine, N.E. Mayo, A new approach toward gait training in patients with Parkinson's disease, Gait Posture 81 (2020) 14–20, https://doi.org/10.1016/j.gaitpost.2020.06.031.
[264] E. Watanabe, M. Satoh, T. Konno, M. Hirai, T. Yamaguchi, The trans-visible navigator: a see-through neuronavigation system using augmented reality, World Neurosurg. 87 (2016) 399–405, https://doi.org/10.1016/j.wneu.2015.11.084.
[265] J.M. Gonçalves, M.J. Sánchez-Ledesma, P. Ruisoto, M. Jaramillo, J.J. Jiménez, J.A. Juanes, Neuronavigation software to visualize and surgically approach brain structures, in: Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality, New York, NY, USA, Association for Computing Machinery, 2018, pp. 405–409, https://doi.org/10.1145/3284179.3284248.
[266] C.F. Dibble, C.A. Molina, Device profile of the XVision-spine (XVS) augmented-reality surgical navigation system: overview of its safety and efficacy, Expert Rev. Med. Devices 18 (2021) 1–8, https://doi.org/10.1080/17434440.2021.1865795.
[267] É. Léger, J. Reyes, S. Drouin, T. Popa, J.A. Hall, D.L. Collins, M. Kersten-Oertel, MARIN: an open-source mobile augmented reality interactive neuronavigation system, Int. J. Comput. Assist. Radiol. Surg. 15 (2020) 1013–1021, https://doi.org/10.1007/s11548-020-02155-6.
[268] Y. Hou, L. Ma, R. Zhu, X. Chen, J. Zhang, A low-cost iPhone-assisted augmented reality solution for the localization of intracranial lesions, PLoS ONE 11 (2016) e0159185.
[269] L. Bréchet, S.B. Hausmann, R. Mange, B. Herbelin, O. Blanke, A. Serino, Subjective feeling of re-experiencing past events using immersive virtual reality prevents a loss of episodic memory, Brain Behav. 10 (2020) 1–9, https://doi.org/10.1002/brb3.1571.
[270] N. Corriveau Lecavalier, É. Ouellet, B. Boller, S. Belleville, Use of immersive virtual reality to assess episodic memory: a validation study in older adults, Neuropsychol. Rehabil. 30 (2020) 462–480, https://doi.org/10.1080/09602011.2018.1477684.
[271] T.D. Parsons, M. Barnett, Validity of a newly developed measure of memory: feasibility study of the virtual environment grocery store, J. Alzheimer's Dis. 59 (2017) 1227–1235, https://doi.org/10.3233/JAD-170295.
[272] P.E. Wais, M. Arioli, R. Anguera-Singla, A. Gazzaley, Virtual reality video game improves high-fidelity memory in older adults, Sci. Rep. 11 (2021) 1–14, https://doi.org/10.1038/s41598-021-82109-3.
[273] G.D. Clemenson, C.E.L. Stark, Virtual environmental enrichment through video games improves hippocampal-associated memory, J. Neurosci. 35 (2015), https://doi.org/10.1523/JNEUROSCI.2580-15.2015.
[274] P. Mateos-Aparicio, A. Rodríguez-Moreno, The impact of studying brain plasticity, Front. Cell. Neurosci. 13 (2019), https://doi.org/10.3389/fncel.2019.00066.
[275] S.H. You, S.H. Jang, Y.H. Kim, Y.H. Kwon, I. Barrow, M. Hallett, Cortical reorganization induced by virtual reality therapy in a child with hemiparetic cerebral palsy, Dev. Med. Child Neurol. 47 (2005) 628–635, https://doi.org/10.1017/S0012162205001234.
[276] Z.R. Wang, P. Wang, L. Xing, L.P. Mei, J. Zhao, T. Zhang, Leap Motion-based virtual reality training for improving motor functional recovery of upper limbs and neural reorganization in subacute stroke patients, Neural Regen. Res. 12 (2017) 1823–1831, https://doi.org/10.4103/1673-5374.219043.
[277] X. Xiao, Q. Lin, W.L. Lo, Y.R. Mao, X.C. Shi, R.S. Cates, S.F. Zhou, D.F. Huang, L. Li, Cerebral reorganization in subacute stroke survivors after virtual reality-based training: a preliminary study, Behav. Neurol. 2017 (2017), https://doi.org/10.1155/2017/6261479.
[278] D. Anopas, Y. Wongsawat, Virtual reality game for memory skills enhancement based on QEEG, in: BMEiCON 2014 - 7th Biomedical Engineering International Conference, 2014, https://doi.org/10.1109/BMEiCON.2014.7017399.

[279] J. Fernandes, A. Teles, S. Teixeira, An augmented reality-based mobile application facilitates the learning about the spinal cord, Educ. Sci. (Basel) 10 (2020) 1–18, https://doi.org/10.3390/educsci10120376.
[280] V. Souza, A. Maciel, L. Nedel, R. Kopper, K. Loges, E. Schlemmer, The effect of virtual reality on knowledge transfer and retention in collaborative group-based learning for neuroanatomy students, in: Proceedings - 2020 22nd Symposium on Virtual and Augmented Reality, SVR 2020, 2020, pp. 92–101, https://doi.org/10.1109/SVR51698.2020.00028.
[281] S.W. Kim, J.K. Seong, Virtual display of 3D computational human brain using Oculus Rift, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 9748 (2016) 258–265. https://doi.org/10.1007/978-3-319-40406-6_24.
[282] R.A. Kockro, C. Amaxopoulou, T. Killeen, W. Wagner, R. Reisch, E. Schwandt, A. Gutenberg, A. Giese, E. Stofft, A.T. Stadie, Stereoscopic neuroanatomy lectures using a three-dimensional virtual reality environment, Ann. Anat. 201 (2015) 91–98, https://doi.org/10.1016/j.aanat.2015.05.006.
[283] A. Gaggioli, A. Ferscha, G. Riva, S. Dunne, I. Viaud-Delmon, Embodied medicine: what human-computer confluence can offer to health care, in: Human Computer Confluence, De Gruyter Open Poland, 2016, pp. 55–79, https://doi.org/10.1515/9783110471137-004.
[284] T.D. Parsons, Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences, Front. Hum. Neurosci. 9 (2015) 660, https://doi.org/10.3389/fnhum.2015.00660.
[285] T.D. Parsons, A. Gaggioli, G. Riva, Virtual reality for research in social neuroscience, Brain Sci. 7 (2017), https://doi.org/10.3390/brainsci7040042.
[286] P. Cipresso, S. Serino, G. Riva, Psychometric assessment and behavioral experiments using a free virtual reality platform and computational science, BMC Med. Inform. Decis. Mak. 16 (2016) 37, https://doi.org/10.1186/s12911-016-0276-5.
[287] G. Riva, B.K. Wiederhold, F. Mantovani, Neuroscience of virtual reality: from virtual exposure to embodied medicine, Cyberpsychol. Behav. Soc. Netw. 22 (2019) 82–96, https://doi.org/10.1089/cyber.2017.29099.gri.
[288] B. Serrano, R.M. Baños, C. Botella, Virtual reality and stimulation of touch and smell for inducing relaxation: a randomized controlled trial, Comput. Human Behav. 55 (2016) 1–8, https://doi.org/10.1016/j.chb.2015.08.007.
[289] R.J. Adams, B. Hannaford, Control law design for haptic interfaces to virtual reality, IEEE Trans. Control Syst. Technol. 10 (2002) 3–13, https://doi.org/10.1109/87.974333.
[290] P. Gamito, J. Oliveira, D. Morais, P. Rosa, T. Saraiv, NeuAR – A Review of the VR/AR Applications in the Neuroscience Domain, in: Augmented Reality - Some Emerging Application Areas, InTech, 2011, https://doi.org/10.5772/26135.
[291] F. Zhou, H.B.L. Dun, M. Billinghurst, Trends in augmented reality tracking, interaction and display: a review of ten years of ISMAR, in: Proceedings - 7th IEEE International Symposium on Mixed and Augmented Reality 2008 2008, ISMAR, 2008, pp. 193–202, https://doi.org/10.1109/ISMAR.2008.4637362.
[292] C.J. Bohil, B. Alicea, F.A. Biocca, Virtual reality in neuroscience research and therapy, Nat. Rev. Neurosci. 12 (2011) 752–762, https://doi.org/10.1038/nrn3122.
[293] Industry insights into the future of immersive technology 2020 augmented and virtual reality survey report 2 Perkins Coie LLP |, (2020).
[294] Augmented reality (AR) & virtual reality (VR) in healthcare market report, 2025, (2017).
[295] G. Riva, S. Serino, D. Di Lernia, E.F. Pavone, A. Dakanalis, Embodied medicine: mens sana in corpore virtuale sano, Front. Hum. Neurosci. 11 (2017), https://doi.org/10.3389/fnhum.2017.00120.
[296] Johnny Kung, The future of neuroscience and VR - CIFAR, (2019).

Physics of Plasmas

RESEARCH ARTICLE | FEBRUARY 01 2024

Localization of beam generated whistler wave and turbulence generation in reconnection region of magnetopause

Jyoti ; Suresh. C. Sharma ; R. P. Sharma

Check for updates

Phys. Plasmas 31, 022902 (2024)
https://doi.org/10.1063/5.0169397

View Online
Export Citation
CrossMark

21 February 2024 06:14:50



AIP Publishing

Localization of beam generated whistler wave and turbulence generation in reconnection region of magnetopause

Cite as: Phys. Plasmas **31**, 022902 (2024); doi: 10.1063/5.0169397
Submitted: 25 July 2023 · Accepted: 5 January 2024 ·
Published Online: 1 February 2024

View Online Export Citation CrossMark

Jyoti,[1,a] Suresh. C. Sharma,[1,b] and R. P. Sharma[2,c]

AFFILIATIONS

[1]Plasma and Nano-Simulation Laboratory, Department of Applied Physics, Delhi Technological University, Shahbad Daulatpur, Bawana Road, Delhi 110042, India
[2]Plasma Simulation Laboratory, Department of Energy Science and Engineering, Indian Institute of Technology, Delhi, Hauz Khas, New Delhi 110016, India

[a]Electronic mail: jyotirani1506@gmail.com
[b]Author to whom correspondence should be addressed: Suresh321sharma@gmail.com
[c]Electronic mail: rpsiitd@gmail.com

ABSTRACT

Whistler waves have been studied for many years in relation to turbulence and particle heating, and observations show that they are crucial to magnetic reconnection. Recent research has revealed a close relationship between magnetic reconnection and turbulence. The current work investigates the whistler turbulence caused by the energetic electron beam in the magnetic reconnection sites of magnetopause and also due to dynamic evolution of magnetic islands. For this, we develop a model based upon the two-fluid approximation to study whistler dynamics, propagating in the medium with the pre-existing chain of magnetic islands and under the influence of background density perturbation originating from ponderomotive nonlinearity of wave. Dynamics of nonlinear whistler have been solved with pseudo-spectral approach and a finite difference method with a modified predictor–corrector method and a Runge Kutta method for the semianalytical model. In the current research, we study how the nonlinear whistler wave contributes to the significant space phenomenon, i.e., turbulence, localization, and magnetic reconnection. We have also investigated the formation of a current sheet in a magnetopause region of the order of few-electron inertial length. We analyzed the power spectrum at the magnetopause when the system reached a quasi-steady condition. Our new approach to study whistler turbulence by an energetic electron beam at the magnetic reconnection sites has extensive applications to space plasmas, shedding a new light on the study of magnetic reconnection in nature.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0169397

21 February 2024 06:14:50

I. INTRODUCTION

One of the most important plasma phenomena is known as magnetic reconnection, which includes a swift topological switch in the magnetic field and an effective conversion of magnetic energy to the kinetic energy. In laboratory, space, and astrophysical plasmas, magnetic field lines alter their topology and transform magnetic energy into plasma particles by acceleration and heating. Magnetic reconnection may be seen in the generation of solar flares, coronal mass ejections, and the interface of solar winds with the Earth's magnetosphere, and it is considered to originate during the formation of stars.[1–8] The astonishing potential of astrophysical plasmas to establish magnetic structure ultimately leads to the accumulation of magnetic energy inside stressed areas, including current sheets. During the process of magnetic reconnection, which results in a reconfiguration of the magnetic field as well as high-speed flows, thermal heating, and nonthermal particle acceleration, this stored energy is frequently released explosively.[9–15] Also, a review article on the theory of magnetic reconnection has been reported earlier.[16] Recently, the coalescence process of numerous magnetic islands originally created in an extended current sheet has been studied, with a particular emphasis on the scaling of the size of the inner structure of the coalesced islands and the accompanying progressive plasma heating process.[17] The reconnection can be examined *in situ* through space applications and laboratory studies.[18] However, only *in situ* space observations can offer precise data on the fields and particle distribution functions.

Published under an exclusive license by AIP Publishing

In space plasma, turbulence is prevalent and is recognized to play a significant role as it cascades energy from large to smaller scales, ultimately leading to dissipation and particles heating. There is mounting evidence that in magnetized plasmas, magnetic reconnection, and turbulence are closely related.[19–22] The Earth's magnetosheath, magnetopause, and magnetotail are all potential sites for magnetic reconnection.[23,24] Many authors have explored how turbulence affects the rate of reconnection, specifically how turbulence already present might change Sweet–Parker reconnection and how turbulence may form because of reconnection.[19] Forced turbulence experiments indicate that turbulent reconnection is quicker than laminar reconnection, and the reconnection rate rises with increasing turbulence level.[25] Additionally, magnetic reconnection itself causes turbulence, which feeds back on itself. Many previous studies investigated the turbulence described as the nonlinear evolution of field and large-scale islands produced at reconnection zones producing the multiple X–O point.[26] Eastwood *et al.*[27] reported the spacecraft observations of turbulence generated by the magnetic reconnection and fluctuations in electric and magnetic field followed by power law of scaling $-5/3$ and $-8/3$. Dynamic alignment models[28–30] of turbulence tend to anticipate perpendicular spectral indices close to $-3/2$, whereas the original "GS95" model,[31] which does not include dynamic alignment, predicts a spectral index of $-5/3$. Surprisingly, it has yet to be determined quantitatively which of these two types of models is true. The new perspective on reconnection relevant in space and astrophysical contexts, where plasma is generally in a fully turbulent regime is numerically analyzed by simulations of two-dimensional magnetohydrodynamic turbulence reveals the presence of a large number of X-type neutral points where magnetic reconnection occurs.[32] Numerical and analytical studies suggest that when magnetic reconnection events occur within the turbulent plasma, they can disrupt these idealized scaling laws.[30,33] Franci *et al.*[34] provide numerical evidence that magnetic reconnection can act as a driver for the onset of the sub-ion turbulent cascade. However, Adhikari *et al.*[35] gives a new direction on relation of reconnection and turbulence, they examine the properties of energy transfer in reconnection and turbulence and found that the energy transfer in both simulations is found to be structurally very similar, with most of the energy transfer occurring through incompressive channels. This provides evidence that reconnection dynamics involves energy transfer analogous to standard turbulence. Recently, Adhikari *et al.*[36] extend the turbulence diagnostics to the case of guide field reconnection and determine how the degree of magnetic shear modifies the turbulence-like properties of reconnection. Although turbulence and reconnection appear to be closely related and it is known that turbulence impacts reconnection, the mechanics of their interaction is still not entirely established.

Different wave modes are detected in earth's reconnection region, like whistler wave,[37–43] electron cyclotron waves,[38,44] kinetic Alfvén waves,[45,46] upper hybrid waves,[47,48] lower hybrid waves,[49,50] and electrostatic solitary waves.[51–53] Several waves, including whistler-mode, have been seen at the magnetic reconnection regions generated due to the energetic electron beam.[41,54–56] Recently, whistler-mode has been reported at Mars[57] and their generation mechanism in magnetic reconnection is based on *in situ* measurements from Mars Atmosphere and Volatile Evolution Mission (MAVEN). Cerri *et al.* also observed that large-scale turbulent motions establish a scaling of $-5/3$ spectrum at $k\lambda_i < 1$ and, at the same time, feed the formation of current sheets where magnetic reconnection occurs.[58] They also investigated the role of fast magnetic reconnection in the development of a quasi-steady turbulent state by means of 2D-3V high-resolution Vlasov–Maxwell simulations. In past, Stenflo *et al.*[59] reported the nonlinear coupling of electron whistler mode and magnetosonic waves but without considering the concept of magnetic island and found the relevance to the non-thermal fluctuations in magnetic field which lead to the plasma heating.

The present work related to whistler wave generation and it's development into localized coherent structures in the presence of magnetic islands is motivated by Zhao *et al.*[55] work, in which the coherent structures of whistler wave at magnetic reconnection locations has been reported due to intense electron beams. Using findings from the magnetospheric multiscale (MMS) mission, they studied the whistler waves generated by intense electron beam in the dayside magnetosphere and observed the maximum growth rate of the wave at $k\lambda_e \sim 0.53$ within the frequency band of $0.1\omega_{ce} - 0.5\omega_{ce}$. Here, λ_e is the electron skin-depth, k is the wave number, and ω_{ce} is electron gyrofrequency. Although, at the initial stage, these structures are coherent, but it is expected that at later stage it might be turbulent. To the best of our knowledge, no particular research concerning beam-driven whistler localization and turbulence has been documented in the previously reported work, despite a few works having been described in the framework of beam-driven turbulence[60,61] but without taking the influence of magnetic islands.

The main goals of this study are to illustrate significant developments in understanding the generation of whistler wave due to the intense electron beam at magnetic reconnection as observed by magnetospheric multiscale mission (MMS)[55] and under the influence of magnetic island. In order to achieve this, a three-dimensional model of a beam-driven whistler wave has been developed. In this model, the whistler wave initiates out owing to the beam's energy and grows to a huge amplitude, where nonlinear effects from the ponderomotive force cause the whistler wave to localize, which ultimately causes the turbulent state. To grasp the nonlinear phase of wave development and saturation, we take into account the nonlinear interaction of a low-frequency ion acoustic wave (IAW) and a high-frequency whistler wave, as well as the ponderomotive nonlinearity caused by the whistler wave. Along with this, we are also analyzing the dynamics of nonlinear whistler wave in the presence of magnetic islands. Then, the governing equations for the IAWs and the whistler wave are then formulated as coupled nonlinear equations and numerically solved with the pseudo-spectral method for spatial integration and a modified version of predictor–corrector method for temporal integration.

The content of this paper is organized as follows: in Sec. II, dynamics of 3D whistler wave generated by intense electron beam and dynamics of ion acoustic wave are discussed. Numerical simulation methods and results obtained from simulation are discussed in Sec. III. In order to understand the basic physics behind the localization, we have also studied the semianalytical model in Sec. IV. Summary and conclusion of our current research work is given in Sec. V.

II. BEAM-DRIVEN WHISTLER WAVE DYNAMICS

The dynamical equation for a 3D whistler wave with an ambient magnetic field along the z-axis in a magnetized plasma propagating in the x, y, and z planes with a wave vector $\vec{k} = k_x\hat{x} + k_y\hat{y} + k_z\hat{z}$ is derived utilizing two fluid models. The governing dynamical equation of 3D whistler wave in terms of electric field can be written as (the detailed derivation has been given as the Appendix)

21 February 2024 06:14:50

Published under an exclusive license by AIP Publishing

$$\frac{\partial^2 E_z}{\partial t^2}+\frac{\partial^6 E_z}{\partial t^2\partial x^4}\lambda_e^4+\frac{\partial^6 E_z}{\partial t^2\partial y^4}\lambda_e^4+\frac{\partial^6 E_z}{\partial t^2\partial z^4}\lambda_e^4+2\lambda_e^4\frac{\partial^4 E_z}{\partial t^2\partial z^2}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)-2\lambda_e^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\frac{\partial E_z}{\partial t^2}$$
$$+\lambda_i^2 v_A^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\frac{\partial^2 E_z}{\partial z^2}+\lambda_i^2 v_A^2\frac{\partial^4 E_z}{\partial z^4}+\frac{\lambda_i^2 v_A^2 B_{0y}^2}{B_{0z}^2}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\frac{\partial^2 E_z}{\partial y^2}+\frac{\lambda_i^2 v_A^2 B_{0y}^2}{B_{0z}^2}\frac{\partial^4 E_z}{\partial y^2\partial z^2}$$
$$+\frac{2\lambda_i^2 v_A^2 B_{0y}}{B_{0z}}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\frac{\partial^2 E_z}{\partial y\partial z}+\frac{2\lambda_i^2 v_A^2 B_{0y}}{B_{0z}}\frac{\partial^4 E_z}{\partial y\partial z^3}=0, \tag{1}$$

where $\omega_{ce}\left(=\frac{eB_0}{m_e c}\right)$ is the electron gyrofrequency of wave, $v_A\left(=\sqrt{\frac{B_0^2}{4\pi n_0 m_i}}\right)$ is Alfvén waves's speed, n_0 denotes the background number density, c denotes the velocity of light, and $\lambda_i\left(=\sqrt{\frac{c^2 m_i}{4\pi n_0 e^2}}\right)$ is the collisionless ion skin depth, $\lambda_e\left(=\sqrt{\frac{c^2 m_e}{4\pi n_0 e^2}}\right)$ is the collisionless electron skin depth. Additionally, it is reported that beam-driven growth rate is taken into account then, the aforementioned equation becomes $\frac{d}{dt}\rightarrow\left(\frac{d}{dt}+\gamma\right)$(see the Appendix),

$$\left(\frac{\partial^2}{\partial t^2}+2\gamma\frac{\partial}{\partial t}\right)E_z+\frac{\partial^6 E_z}{\partial t^2\partial x^4}\lambda_e^4+\frac{\partial^6 E_z}{\partial t^2\partial y^4}\lambda_e^4+\frac{\partial^6 E_z}{\partial t^2\partial z^6}\lambda_e^4+2\lambda_e^4\frac{\partial^4 E_z}{\partial t^2\partial z^2}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)$$
$$-2\lambda_e^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\frac{\partial^2 E_z}{\partial t^2}+\lambda_i^2 v_A^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\frac{\partial^2 E_z}{\partial z^2}+\lambda_i^2 v_A^2\frac{\partial^4 E_z}{\partial z^4}$$
$$+\frac{\lambda_i^2 v_A^2 B_{0y}^2}{B_{0z}^2}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\frac{\partial^2 E_z}{\partial y^2}+\frac{\lambda_i^2 v_A^2 B_{0y}^2}{B_{0z}^2}\frac{\partial^4 E_z}{\partial y^2\partial z^2}+\frac{2\lambda_i^2 v_A^2 B_{0y}}{B_{0z}}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\frac{\partial^2 E_z}{\partial y\partial z}+\frac{2\lambda_i^2 v_A^2 B_{0y}}{B_{0z}}\frac{\partial^4 E_z}{\partial y\partial z^3}=0. \tag{2}$$

For Eq. (2), the following envelope solution is assumed:

$$\tilde{E}_z=E_z(x,y,z,t)e^{i(k_{0x}\hat{x}+k_{0y}\hat{y}+k_{0z}\hat{z}-\omega_0 t)}. \tag{3}$$

By taking into account the impact of a whistler wave on the background density and relieving the aforementioned solution into Eq. (2), we can obtain the dynamical equation of a nonlinear whistler wave propagating through plasma

$$2i\omega_0(1+\lambda_e^4k_0^4+2\lambda_e^2k_0^2+2\lambda_e^4k_{0x}^2k_{0z}^2+2\lambda_e^4k_{0y}^2k_{0z}^2)\frac{\partial E_z}{\partial t}+(4\omega_0^2\lambda_e^4k_{0x}^2)\frac{\partial^2 E_z}{\partial x^2}+(4\omega_0^2\lambda_e^4k_{0y}^2)\frac{\partial^2 E_z}{\partial y^2}+(4\omega_0^2\lambda_e^4k_{0z}^2-4\lambda_i^2v_A^2k_{0z}^2)\frac{\partial^2 E_z}{\partial z^2}$$
$$+2ik_{0x}(2\omega_0^2k_{0x}^2\lambda_e^4+2\omega_0^2k_{0z}^2\lambda_e^4+2\omega_0^2\lambda_e^2+k_{0z}^2\lambda_i^2v_A^2)\frac{\partial E_z}{\partial x}+2ik_{0y}(2\omega_0^2k_{0y}^2\lambda_e^4+2\omega_0^2k_{0z}^2\lambda_e^4+2\omega_0^2\lambda_e^2-k_{0z}^2\lambda_i^2v_A^2)\frac{\partial E_z}{\partial y}$$
$$+2ik_{0z}(2\omega_0^2k_{0z}^2\lambda_e^4+2\omega_0^2k_{0x}^2\lambda_e^4+2\omega_0^2k_{0y}^2\lambda_e^4+2\omega_0^2\lambda_e^2-k_{0y}^2\lambda_i^2v_A^2-2k_{0z}^2\lambda_i^2v_A^2)\frac{\partial E_z}{\partial z}$$
$$+(4\lambda_i^2v_A^2k_{0x}k_{0z}-8\omega_0^2\lambda_e^4k_{0z}k_{0x})\frac{\partial^2 E_z}{\partial x\partial z}+(4\lambda_i^2v_A^2-8\omega_0^2\lambda_e^4)k_{0y}k_{0z}\frac{\partial^2 E_z}{\partial y\partial z}+\frac{\lambda_i^2v_A^2B_{0y}^2}{B_{0z}^2}k^2k_{0y}^2E_z+\frac{2\lambda_i^2v_A^2B_{0y}}{B_{0z}}k^2k_{0y}k_{0z}E_z$$
$$-(\lambda_i^2v_A^2k^2k_{0z}^2)\frac{\delta n}{n_0}E_z+2i\gamma' E_z=0, \tag{4}$$

where k_{0x}, k_{0y}, and k_{0z} are wave vector components of whistler related to the background magnetic field are given as $k_0^2=k_{0x}^2+k_{0y}^2+k_{0z}^2$. Here, n' denotes the modified density, i.e., $n'=n_0+\delta n$, and $\frac{\delta n}{n_0}$ denotes the pertubation in plasma density due to non-linear whistler wave. we have used the relation of the magnetic field and vector potential $\left(\frac{B_{0y}}{B_{0z}}=\frac{\delta A_z}{A_0}\right)$. ($\delta A_z$, is the vector potential associated with magnetic field.)

Equation (4) in normalized dimensionless form is given as

$$ic_1\frac{\partial E_z}{\partial t}+c_2\frac{\partial^2 E_z}{\partial x^2}+c_3\frac{\partial^2 E_z}{\partial y^2}+c_4\frac{\partial^2 E_z}{\partial z^2}+ic_5\frac{\partial E_z}{\partial x}+ic_6\frac{\partial E_z}{\partial y}+ic_7\frac{\partial E_z}{\partial z}$$
$$+c_8\frac{\partial^2 E_z}{\partial x\partial z}+c_9\frac{\partial^2 E_z}{\partial y\partial z}+c_{10}\left(\frac{\delta A_z}{A_0}\right)^2E_z+c_{11}\left(\frac{\delta A_z}{A_0}\right)E_z-\frac{\delta n}{n_0}E_z+2i\gamma' E_z=0, \tag{5}$$

where Eq. (5) represents the whistler wave's normalized dynamical equation, which causes turbulence. The normalizing parameters are

$$t_n=1/\omega_0,$$

21 February 2024 06:14:50

Published under an exclusive license by AIP Publishing

$$x_n = \frac{k_{0x}(2\omega_0^2 k_{0x}^2\lambda_e^4 + 2\omega_0^2 k_{0z}^2\lambda_e^4 + 2\omega_0^2\lambda_e^2 + k_{0z}^2\lambda_i^2 v_A^2)}{\omega_0(1+\lambda_e^4 k_0^4 + 2\lambda_e^2 k_0^2 + 2\lambda_e^4 k_{0x}^2 k_{0z}^2 + 2\lambda_e^4 k_{0y}^2 k_{0z}^2)},$$

$$z_n = \frac{2i(-2\omega_0^2 k_{0x}^3\lambda_e^4 + 2\omega_0^2 k_{0x}k_{0z}^2\lambda_e^4 + 2\omega_0^2\lambda_e^2 k_{0x} + k_{0x}k_{0z}^2\lambda_i^2 v_A^2 + 2k_{0z}^3\lambda_i^2 v_A^2)}{2\omega_0(1+\lambda_e^4 k_{0x}^4 + \lambda_e^4 k_{0z}^4 + 2\lambda_e^2 k_{0x}^2 + 2\lambda_e^2 k_{0z}^2)}.$$

In this section, the beam instability that generates the whistler wave will be described. The growth rate of whistler wave generated by an energetic electron beam, i.e., is included phenomenologically and its value corresponding with the data published by Zhao *et al.*[55] will be utilized in simulations here. The field perturbation due to magnetic islands field is incorporated in our model,[62] $\frac{\delta A_z}{A_0} = (-\frac{x^2}{2} + b_0\cos(ky))$. So, using the profile of field perturbation Eq. (5) modifies as

$$ic_1\frac{\partial E_z}{\partial t} + c_2\frac{\partial^2 E_z}{\partial x^2} + c_3\frac{\partial^2 E_z}{\partial y^2} + c_4\frac{\partial^2 E_z}{\partial z^2} + ic_5\frac{\partial E_z}{\partial x} + ic_6\frac{\partial E_z}{\partial y} + ic_7\frac{\partial E_z}{\partial z} + c_8\frac{\partial^2 E_z}{\partial x\partial z} + c_9\frac{\partial^2 E_z}{\partial y\partial z} + c_{10}\left(-\frac{x^2}{2} + b_0\cos(ky)\right)^2 E_z + c_{11}\left(-\frac{x^2}{2} + b_0\cos(ky)\right)E_z - \frac{\delta n}{n_0}E_z + 2i\gamma' E_z = 0, \tag{6}$$

where

$$c_1 = \frac{2\omega_0(1+\lambda_e^4 k_0^4 + 2\lambda_e^2 k_0^2 + 2\lambda_e^4 k_{0x}^2 k_{0z}^2 + 2\lambda_e^4 k_{0y}^2 k_{0z}^2)}{(\lambda_i^2 v_A^2 k^2 k_{0z}^2)}, \quad c_2 = \frac{2k_{0x}(2\omega_0^2\lambda_e^4 k_{0x}^2 + 2\omega_0^2\lambda_e^4 k_{0z}^2 + 2\omega_0^2\lambda_e^2 - \lambda_i^2 k_{0z}^2 v_A^2)}{(\lambda_i^2 v_A^2 k^2 k_{0z}^2)},$$

$$c_3 = \frac{2k_{0y}(2\omega_0^2\lambda_e^4 k_{0y}^2 + 2\omega_0^2\lambda_e^4 k_{0z}^2 + 2\omega_0^2\lambda_e^2 - \lambda_i^2 k_{0z}^2 v_A^2)}{(\lambda_i^2 v_A^2 k^2 k_{0z}^2)}, \quad c_4 = \frac{2k_{0z}(2\omega_0^2\lambda_e^2 k_{0z}^2 + 2\omega_0^2\lambda_e^2 k_{0x}^2 + 2\omega_0^2\lambda_e^2 k_{0y}^2 + 2\omega_0^2\lambda_e^2 - \lambda_i^2 v_A^2 k_{0y}^2 - 2\lambda_i^2 v_A^2 k_{0z}^2)}{\lambda_i^2 v_A^2 k^2 k_{0z}^2},$$

$$c_5 = \frac{4\omega_0^2 k_{0x}^2\lambda_e^4}{\lambda_i^2 v_A^2 k_0^2 k_{0z}^2}, \quad c_6 = \frac{4\omega_0^2 k_{0y}^2\lambda_e^4}{\lambda_i^2 v_A^2 k_0^2 k_{0z}^2}, \quad c_7 = \frac{(4\omega_0^2 k_{0z}^2\lambda_e^4 - 4\lambda_i^2 v_A^2 k_{0z}^2)}{\lambda_i^2 v_A^2 k_0^2 k_{0z}^2}. \; b_0 = 0.5, \quad c_8 = \frac{(8\omega_0^2\lambda_e^4 - 4\lambda_i^2 v_A^2)k_{0z}k_{0x}}{\lambda_i^2 v_A^2 k_0^2 k_{0z}^2},$$

$$c_9 = \frac{(8\omega_0^2\lambda_e^4 - 4\lambda_i^2 v_A^2)k_{0z}k_{0y}}{\lambda_i^2 v_A^2 k_0^2 k_{0z}^2}, \quad c_{10} = \frac{k_{0y}^2}{k_{0z}^2}, \quad \text{and} \quad c_{11} = \frac{2k_{0y}}{k_{0z}},$$

and $\gamma' = \frac{\gamma}{\omega_0}$ is the normalized growth rate. $\gamma(= 0.5083\omega_{ce})$ is the growth rate of beam-driven whistler waves.

A. Ion acoustic wave (IAW) dynamics

Consider a low-frequency ion-acoustic wave propagating parallel to the background magnetic field B_0 along the z-axis, i.e., $\vec{B}_0 = B_0\hat{z}$, $\vec{k} = k_z\hat{z}$. To establish the IAW dynamics, the following key equations are used:

The equation of motion

$$m_j\frac{\partial\vec{v}_j}{\partial t} = q_j\vec{E} + \frac{q_j}{c}(\vec{v}_j\times\vec{B}_0) - T_j\vec{\nabla}\frac{n_j}{n_0} + \vec{F}_j. \tag{7}$$

The continuity equation

$$\frac{\partial n_j}{\partial t} + \vec{\nabla}.(n_j\vec{v}_j) = 0, \tag{8}$$

where F_j denotes the ponderomotive force of the high-frequency whistler wave and v_j denotes the velocity of the species (j stands for an electron's velocity "e" and "i" for an ion's velocity) of IAW.

When we assume the quasi-neutrality along with the response of the electrons and use the parallel element of the linearized version of Eq. (7) for the ions, we obtain

$$\frac{\partial v_{iz}}{\partial t} = -\frac{c_s^2}{n_0}\frac{\partial n_i}{\partial z} + \left(\frac{F_{iz}+F_{ez}}{m_i}\right), \tag{9}$$

where $c_s(=\frac{k_B(T_e+T_i)}{m_i})^{1/2}$ denotes the IAW's, k_B denotes the Boltzmann Constant. T_i, T_e are the temperature of ion and electron, respectively. F_{ez} (F_{iz}) denotes the constituent of ponderomotive force in z-direction formed by whistler wave.

Differentiating Eq. (8) with respective to time again, we get

$$\frac{\partial^2 n_i}{\partial t^2} + n_0\frac{\partial^2 v_{iz}}{\partial t\partial z} = 0. \tag{10}$$

Combining Eqs. (9) and (10), we obtain the nonlinear dynamical equation of IAW

$$\frac{\partial^2 n}{\partial t^2} = -n_0\frac{\partial}{\partial z}\left[-c_s^2\frac{\partial}{\partial z}\left(\frac{n}{n_0}\right) + \frac{F_{ez}+F_{iz}}{m_i}\right]. \tag{11}$$

Equation (11) can be redrafted as

$$\left(\frac{\partial^2 n}{\partial t^2} - c_s^2\frac{\partial^2}{\partial z^2}\right)\left(\frac{n}{n_0}\right) = -\frac{\partial}{\partial z}\left[\frac{F_{ez}+F_{iz}}{m_i}\right]. \tag{12}$$

Now, the Whistler Wave's Ponderomotive Force is stated as

$$\vec{F}_j = -m_j(\vec{v}_j^*.\vec{\nabla})\vec{v}_j + \frac{q_j}{c}(\vec{v}_j^*\times B_w). \tag{13}$$

Here, the charged species (ions and electrons) are indicated by "j." Thus, q_j, v_j, and m_j denotes the charge, velocity, and mass of ion and electrons, respectively. Magnetic field due to the whistler wave is denoted by B_w and c denotes the speed of light.

21 February 2024 06:14:50

Published under an exclusive license by AIP Publishing

The whistler wave's velocity component owing to electrons may be expressed as

$$v_{e1x} = \frac{-ie\omega_0}{m_e(\omega_0^2 - \omega_{ce}^2)}E_x - \frac{e\omega_{ce}}{m_e(\omega_0^2 - \omega_{ce}^2)}E_y, \tag{14}$$

$$v_{e1y} = \frac{e\omega_{ce}}{m_e(\omega_0^2 - \omega_{ce}^2)}E_x - \frac{ie\omega_0}{m_e(\omega_0^2 - \omega_{ce}^2)}E_y, \tag{15}$$

$$v_{e1z} = \frac{eE_z}{im_e\omega_0}. \tag{16}$$

In a similar manner, the velocity element of the whistler wave generated by an ion may be expressed as

$$v_{i1x} = \frac{ie\omega_0}{m_i(\omega_0^2 - \omega_{ci}^2)}E_x - \frac{e\omega_{ci}}{m_i(\omega_0^2 - \omega_{ci}^2)}E_y, \tag{17}$$

$$v_{i1y} = \frac{e\omega_{ci}}{m_i(\omega_0^2 - \omega_{ci}^2)}E_x + \frac{ie\omega_0}{m_i(\omega_0^2 - \omega_{ci}^2)}E_y, \tag{18}$$

$$v_{i1z} = \frac{-eE_z}{im_i\omega_0}. \tag{19}$$

Relieving the value of the whistler wave velocity components in Eq. (13) to obtain the ponderomotive force components[63] related to whistler waves

$$F_{ez} = \frac{e^2}{2m_ec^2}\frac{\partial|E_z|^2}{\partial z} \quad \text{and} \quad F_{iz} = \frac{e^2}{2m_ic^2}\frac{\partial|E_z|^2}{\partial z}.$$

By substituting the value of F_{iz} and F_{ez} in Eq. (12), one may derive the dynamical equation of IAW,

$$\left(\frac{\partial^2}{\partial t^2} - c_s^2\frac{\partial^2}{\partial z^2}\right)\left(\frac{n}{n_0}\right) = -\frac{\partial}{\partial z}\left[\frac{1}{m_i}\left(\frac{e^2}{2m_ec^2} + \frac{e^2}{2m_ic^2}\right)\right]\frac{\partial^2}{\partial z^2}|E_z|^2. \tag{20}$$

The above-mentioned Eq. (20) represents the non-linear dynamics of IAW. By applying the normalized parameter just like in dynamics of whistler wave, one may establish the normalized dimensionless equation of IAW

$$\left(\frac{\partial^2}{\partial t^2} - c_{12}\frac{\partial^2}{\partial z^2}\right)\left(\frac{n}{n_0}\right) = -c_{13}\frac{\partial^2}{\partial z^2}|E_z|^2, \tag{21}$$

where $c_{12} = c_s^2\frac{t_n^2}{z_n^2}$. and $c_{13} = -\left[\frac{1}{m_i}\left(\frac{e^2}{2m_ec^2} + \frac{e^2}{2m_ic^2}\right)\right]\frac{t_n^2}{z_n^2}$.

Equation (21) simplifies to when we assume the adiabatic response

$$\frac{n}{n_0} = \frac{c_{13}}{c_{12}}|A_z|^2. \tag{22}$$

Using the normalized electric field, we obtained $E_n = \left[\frac{2m_e m_i^2 c^2 c_s^2}{n_0 e^2(m_e + m_i)}\right]^{\frac{1}{2}}$.

Now Eq. (6) becomes

$$\begin{aligned} &ic_1\frac{\partial E_z}{\partial t} + c_2\frac{\partial^2 E_z}{\partial x^2} + c_3\frac{\partial^2 E_z}{\partial y^2} + c_4\frac{\partial^2 E_z}{\partial z^2} + ic_5\frac{\partial E_z}{\partial x} \\ &+ ic_6\frac{\partial E_z}{\partial y} + ic_7\frac{\partial E_z}{\partial z} + c_8\frac{\partial^2 E_z}{\partial x\partial z} + c_9\frac{\partial^2 E_z}{\partial y\partial z} \\ &+ c_{10}\left(-\frac{x^2}{2} + b_0\cos(ky)\right)^2 E_z + c_{11}\left(-\frac{x^2}{2} + b_0\cos(ky)\right)E_z \\ &- \frac{c_{13}}{c_{12}}|A_z|^2E_z + 2i\gamma' E_z = 0. \end{aligned} \tag{23}$$

TABLE I. Testing of invariants with time steps with Grid points 2048 (this step size is also used for the finite difference method to monitor the invariants of NLS equation to desired accuracy).

Time steps (dt)	Plasmon number (N)
10^{-5}	9.999 999 97
10^{-6}	9.999 999 996
10^{-7}	1.000 000 000 07
10^{-8}	1.000 000 000 07

The values of numerical constants are given as follows:

$n_0 = 21\,\text{cm}^{-3}$, $B_0 = 45\,\text{nT}$, $T_e = 42\,\text{eV}$, $T_I = 235\,\text{eV}$, $\omega_{pe} = 2.5 \times 10^5\text{rad/s}$, $\omega_{ce} = 7.9 \times 10^3\text{rad/s}$, $\omega_{ci} = 4.31\text{rad/s}$, $v_A = 2.14 \times 10^7\text{cm/s}$, $c_s = 1.6 \times 10^7\text{cm/s}$, $\lambda_i = 4.9 \times 10^6\text{cm}$, $\lambda_e = 1.16 \times 10^5\text{cm}$, $E_n = 1.5 \times 10^3 G\,\text{cm}$, $x_n = y_n = 1.02 \times 10^6\text{cm}$, $z_n = 1.5 \times 10^6\,\text{cm}$, $t_n = 0.015\,\text{s}$, $b_0 = 0.5$.

$c_1 = 0.0155$, $c_2 = c_3 = 1.02 \times 10^6$, $c_4 = 1.5 \times 10^6$, $c_5 = c_6 = 2.5 \times 10^{10}$, $c_7 = -2.0 \times 10^{11}$, $c_8 = c_9 = -7.2 \times 10^{11}$, $c_{10} = 12.41$, $c_{11} = 7.04$, $c_{12} = 1.6 \times 10^5$, and $c_{13} = 3.9 \times 10^3$.

III. NUMERICAL SIMULATION AND RESULT DISCUSSION

To solve the dynamics of nonlinear whistler wave, numerical simulation has been performed. First, the accuracy of code has been checked by studying the algorithm of nonlinear Schr ö dinger (NLS) equation and consistency of plasmon number, $N = \sum_k |E_k|^2$ up to sixth decimal number into two steps

(i) The testing of invariants plasmon number $N = \sum_k |E_k|^2$ with time steps and grid points is given below (NLS) in Tables I and II.

(ii) Modified for present case and applied to magnetopause parameters below:
Then, the pseudo-spectral method and finite difference method with the step size of $dt = 5 \times 10^{-5}$ have been used in the periodic domain of $(10\pi \times 10\pi \times 10\pi)$ with the grid size of $(128 \times 128 \times 128)$. The initial condition imposed for numerical simulation consists of sequential X–O points represents the magnetic reconnection[64] is

$$E_z(x, y) = \cos(2x + 2.3) + \cos(y + 4.1). \tag{24}$$

Figure 1 depicts the spatial evolution of normalized electric field (E_z) over the different times in x–y plane and at fixed z-plane. Furthermore, we hypothesize that a 3D whistler wave originates within an already-existing reconnection point. Field perturbation owing to the whistler wave as well as density modification caused by ponderomotive force nonlinearity are both responsible for the spatial fluctuation of the whistler wave field. At initial level, structures formed are

TABLE II. Testing of invariants with grid points.

Grid points	256	512	1024	2048
N	9.999 999 1	9.999 997	9.999 999 4	9.999 999 8

21 February 2024 06:14:50

Published under an exclusive license by AIP Publishing

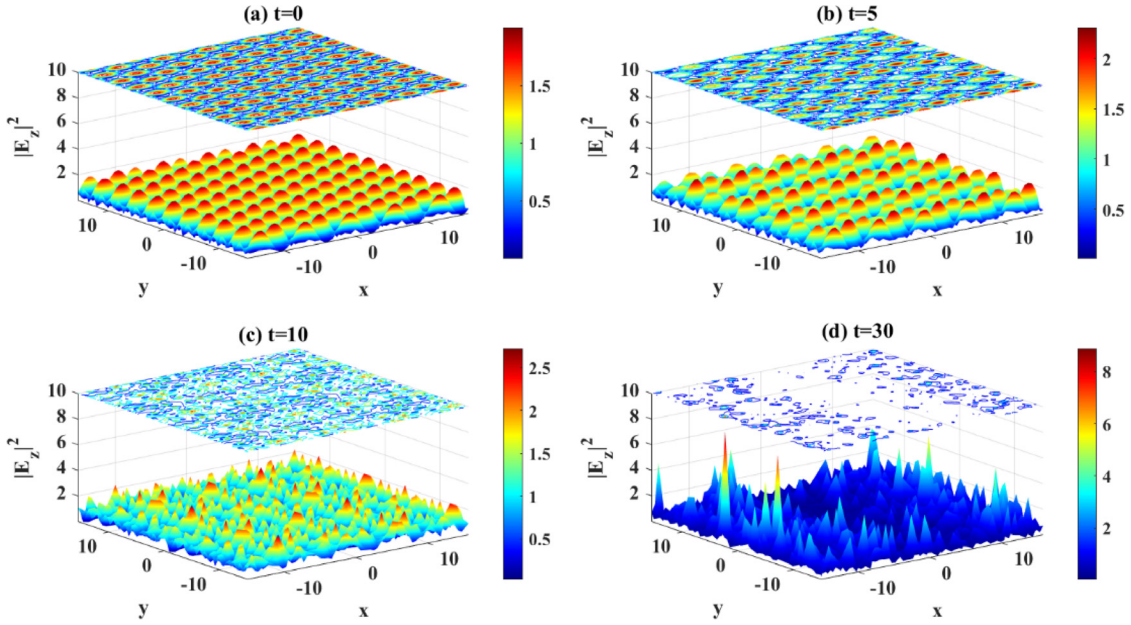


FIG. 1. The projection plot of nonlinear whistler wave in the presence of magnetic islands and perturbation in the background density for different times: (a) t = 0, (b) t = 3, (c) t = 8, and (d) t = 10, for modified nonlinear Schr o dinger equation (in normalized units).

coherent but with increase of the time evolution this will give rise to generation of whistler turbulence as shown in Fig. 1.

Plotting the contour plot of flux lines can be utilized to reveal the existence of magnetic islands created by reconnection. We then examine the spatial evolution of current sheet of whistler wave with t in the presence of reconnection induced magnetic islands as well as background density fluctuations. Figure 2 illustrates the contour plot (left panel of Fig. 2), and current density (right panel of Fig. 2) of nonlinear whistler wave in the presence of magnetic islands and perturbation in the background density in x–y plane at different times. The figure clearly illustrates that for the very initial stage, the current is having a smooth or symmetric distribution, as time progresses, we observe

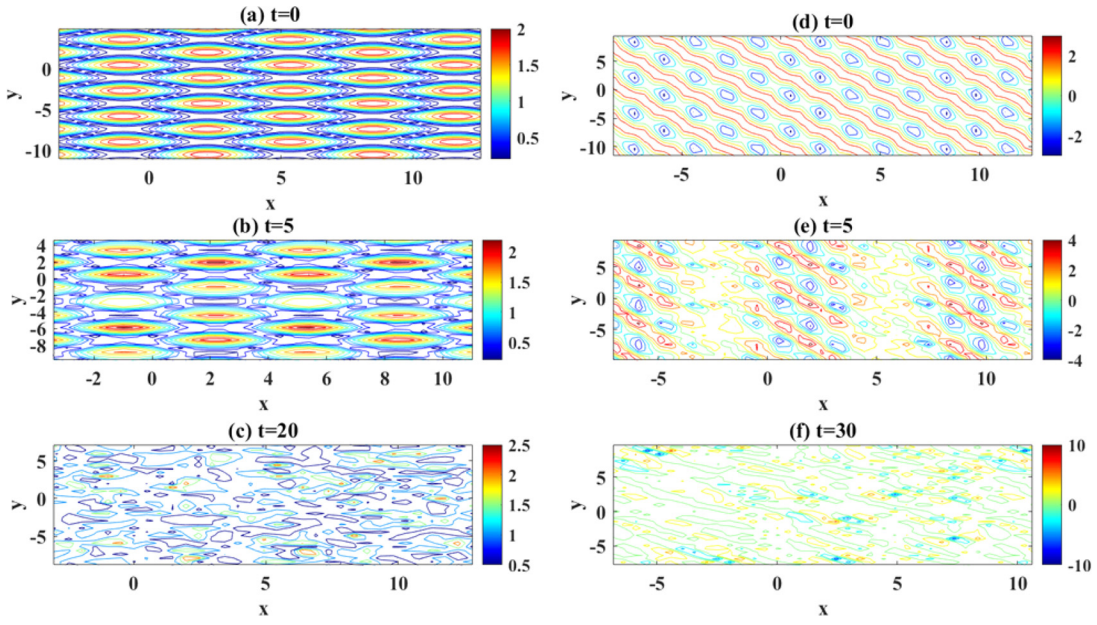


FIG. 2. Contour plot (left panel) and current sheet (right panel) of nonlinear whistler in the presence of magnetic islands and perturbation in the background density for different times (in normalized units).

21 February 2024 06:14:50

Published under an exclusive license by AIP Publishing

that the appropriate X–O structures get deformed, and we eventually reach a wholly chaotic pattern. As a result, these chaotic structures may be responsible for the creation of whistler turbulence.

The quasi-steady state has not yet been reached and would not be for some time since the constant growth rate of the beam-driven mode was taken into account in the present investigation. As a result, addressing power spectra is not relevant, but we would like to improve the present model in future by taking into account the growth rate whistler intensity-dependent (instead of taking constant growth rate). When the whistler wave's amplitude is sufficient, we turn off the growth rate of the beam-driven mode so as to examine the turbulent spectra. So, to examine the ensemble averaged power spectrum for quasi-steady state, we write Eq. (23) as follows:

$$ic_7\frac{\partial E_z}{\partial z}+c_2\frac{\partial^2 E_z}{\partial x^2}+c_3\frac{\partial^2 E_z}{\partial y^2}+ic_5\frac{\partial E_z}{\partial x}+ic_6\frac{\partial E_z}{\partial y}+c_8\frac{\partial^2 E_z}{\partial x\partial z}+c_9\frac{\partial^2 E_z}{\partial y\partial z}-\frac{c_{13}}{c_{12}}|A_z|^2E_z=0. \quad (25)$$

Simulating Eq. (25) for grid size (512×512) and periodic domain is $(10\pi\times10\pi)$ by imposing the initial condition of Biskamp and Welter[64] and then we investigated the ensemble averaged spectra in the presence of magnetic islands (anisotropic current sheet formation also) as shown in Fig. 3(a). It shows the variation of $|E_k|^2$ with k at z = 43–50, i.e., there are seven spectra. We just give the scaling of reference lines (red and green scaling lines for −1.2 and −7/3 scaling) and actual power spectra is given by solid curve line. Due to the presence of magnetic islands it slightly different from Kolmogorov scaling of −5/3. However, in the absence of magnetic islands, power spectrum follows the Kolmogorov scaling in the inertial range that has scaling of −5/3 and after that spectrum is steeper, shown in Fig. 3(b) as we have done in our earlier work.[61] It is to be mentioned here that scaling is different from −5/3 when magnetic islands are present. This is most likely due to the fact that magnetic reconnection possibly modifies the classical Kolmogorov cascade.[33,65] This happens since the anisotropic, current-sheet become the sites of magnetic reconnection (magnetic islands) before the formal Kolmogorov dissipation scale is reached.

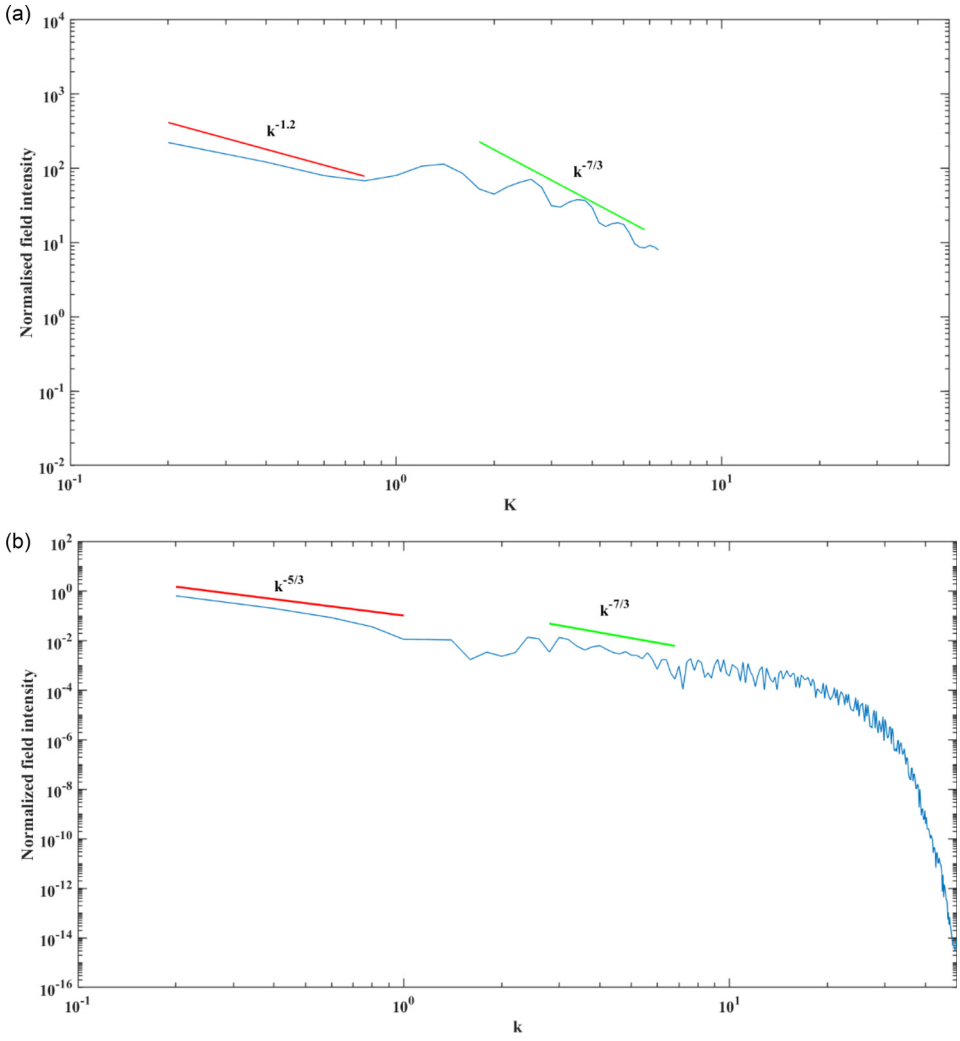


FIG. 3. (a) Ensemble averaged power spectra of Normalized field intensity of nonlinear whistler wave $|E_k|^2$ against normalized wave vector k for modified nonlinear Schrödinger equation (in normalized units) in the presence of magnetic island effects. (b) Ensemble averaged power spectra of Normalized field intensity of nonlinear whistler wave $|E_k|^2$ against normalized wave vector k for modified nonlinear Schr ö dinger equation (in normalized units) in the absence of magnetic island effects.

21 February 2024 06:14:50

Published under an exclusive license by AIP Publishing

As a consequence of the wave's dispersive features, this power spectrum's steepening at smaller scales signifies energy transfer from larger to smaller length scales. This specific spectral index will promote the emergence of thermal tails of energetic particles which causes the acceleration of the particles. The repetitive interaction between charged particles and localized structures can lead to energization using the Fokker–Planck equation (also called second order Fermi acceleration).[66,67] In this method, the repetitive interaction between charged particles and magnetic clouds generated by turbulence lead to the acceleration of charged particles. There are various mechanism have been used to accomplish the study of acceleration of charged particles.[68] The current study presents the localized structures of nonlinear whistler wave based on the nonlinear coupling and turbulence generation. In this chaotic scenario, charged particles gain energy from localized structures by interacting repetitively, leading to particles acceleration which is similar to Fermi acceleration mechanism. To describe the repetitive interaction of charged particles with the localized structures, a quasi-linear diffusion equation is used, and it is provided by the following formula:[69]

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial v}\left\{D(v)\frac{\partial f}{\partial v}\right\}.$$

Here, $D(v)$ denotes the velocity space diffusion coefficient, and $f(t, v)$ denotes the velocity distribution function. The velocity distribution function can be independent of time if the observation time is large as compared to the ponderomotive nonlinearity setting up time. In this context, the relation between spectral index η, and distribution function $f(v)$[70,71] is given as follows:

$$f(v) \propto v^{2-\eta}. \tag{26}$$

In our current analysis $\eta \sim 2.3$; therefore, $f(v) \sim v^{-0.3}$, which leads to the formation of thermal tail of energetic electrons.

IV. SEMIANALYTICAL MODEL

In order to understand the basic physics behind the localization of whistler wave we have investigated the semianalytical model, also by using this model we can estimate the critical size of coherent structures. Now, by modifying Eq. (4), we obtain following equation:

$$\begin{aligned}
&i\frac{2}{k_\parallel}\frac{\partial E_z}{\partial z} + \frac{(4\omega_0^2\lambda_e^4k_\perp^2)}{(2\omega_0^2k_\parallel^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2\lambda_e^2 - k_\perp^2\lambda_i^2v_A^2 - 2k_\parallel^2\lambda_i^2v_A^2)k_\parallel^2}\frac{\partial^2 E_z}{\partial x^2}\\
&+ \frac{(4\omega_0^2\lambda_e^4k_\perp^2)}{(2\omega_0^2k_\parallel^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2\lambda_e^2 - k_\perp^2\lambda_i^2v_A^2 - 2k_\parallel^2\lambda_i^2v_A^2)k_\parallel^2}\frac{\partial^2 E_z}{\partial y^2}\\
&+ \frac{\lambda_i^2v_A^2k^2k_\perp^2}{(2\omega_0^2k_\parallel^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2\lambda_e^2 - k_\perp^2\lambda_i^2v_A^2 - 2k_\parallel^2\lambda_i^2v_A^2)k_\parallel^2}\left(\frac{\delta A_z}{A_0}\right)^2 E_z\\
&+ \frac{2\lambda_i^2v_A^2k^2k_{0y}k_\parallel}{(2\omega_0^2k_\parallel^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2\lambda_e^2 - k_\perp^2\lambda_i^2v_A^2 - 2k_\parallel^2\lambda_i^2v_A^2)k_\parallel}\left(\frac{\delta A_z}{A_0}\right)E_z\\
&- \frac{(\lambda_i^2v_A^2k^2)}{(2\omega_0^2k_\parallel^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2\lambda_e^2 - k_\perp^2\lambda_i^2v_A^2 - 2k_\parallel^2\lambda_i^2v_A^2)}\frac{\delta n}{n_0}E_z + E_z\exp(k_iz) = 0.
\end{aligned} \tag{27}$$

Or we can rewrite above equation in simplified equation as

$$i\frac{2}{k_\parallel}\frac{\partial E_z}{\partial z} + \theta_1\lambda_e^2\left(\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2}\right) + \theta_2\left(-\frac{x^2}{2} + A_z\cos(ky)\right)^2 E_z + \theta_3\left(-\frac{x^2}{2} + A_z\cos(ky)\right)E_z - \theta_4\frac{\delta n}{n_0}E_z + E_z\exp(k_iz) = 0. \tag{28}$$

where

$$\theta_1 = \frac{(4\omega_0^2\lambda_e^2k_\perp^2)}{(2\omega_0^2k_\parallel^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2\lambda_e^2 - k_\perp^2\lambda_i^2v_A^2 - 2k_\parallel^2\lambda_i^2v_A^2)k_\parallel^2},$$

$$\theta_2 = \frac{\lambda_i^2v_A^2k^2k_\perp^2}{(2\omega_0^2k_\parallel^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2\lambda_e^2 - k_\perp^2\lambda_i^2v_A^2 - 2k_\parallel^2\lambda_i^2v_A^2)k_\parallel^2},$$

$$\theta_3 = \frac{2\lambda_i^2v_A^2k^2k_\perp}{(2\omega_0^2k_\parallel^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2\lambda_e^2 - k_\perp^2\lambda_i^2v_A^2 - 2k_\parallel^2\lambda_i^2v_A^2)k_\parallel},$$

$$\theta_4 = \frac{(\lambda_i^2v_A^2k^2)}{(2\omega_0^2k_\parallel^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2k_\perp^2\lambda_e^4 + 2\omega_0^2\lambda_e^2 - k_\perp^2\lambda_i^2v_A^2 - 2k_\parallel^2\lambda_i^2v_A^2)}.$$

Now we are assuming the envelope solution of electric field $\tilde{E}_z = E_{00}e^{ik_{0z}s(x,y,z)}$, and putting this envelope solution in Eq. (28) we obtained the following equation:

21 February 2024 06:14:50

Published under an exclusive license by AIP Publishing

$$i\frac{2}{k_{0z}}\frac{\partial E_{00}}{\partial z}-2E_{00}\frac{\partial s}{\partial z}+\theta_1\lambda_e^2\left(\frac{\partial^2 E_{00}}{\partial x^2}+\frac{\partial^2 E_{00}}{\partial y^2}\right)+2ik_{\parallel}\theta_1\lambda_e^2\left(\frac{\partial E_{00}}{\partial x}\frac{\partial s}{\partial x}+\frac{\partial E_{00}}{\partial y}\frac{\partial s}{\partial y}\right)-k_{\parallel}^2\theta_1\lambda_e^2\left(\frac{\partial s}{\partial x}+\frac{\partial s}{\partial y}\right)^2E_{00}+ik_{\parallel}\theta_1\lambda_e^2\left(\frac{\partial^2 s}{\partial x^2}+\frac{\partial^2 s}{\partial y^2}\right)E_{00}+\theta_2\left[\left(-\frac{x^2}{2}+b_0\cos(ky)\right)\right]^2E_{00}+\theta_3\left[\left(-\frac{x^2}{2}+b_0\cos(ky)\right)\right]E_{00}-\theta_4\frac{\delta n}{n_0}E_{00}+E_{00}\exp(k_iz)=0. \tag{29}$$

Further separating the real and imaginary part of Eq. (29),

$$-2\frac{\partial s}{\partial z}E_{00}+\theta_1\lambda_e^2\left(\frac{\partial^2 E_{00}}{\partial x^2}+\frac{\partial^2 E_{00}}{\partial y^2}\right)-k_{\parallel}^2\theta_1\lambda_e^2\left(\frac{\partial s}{\partial x}+\frac{\partial s}{\partial y}\right)^2E_{00}+\theta_2\left[\left(-\frac{x^2}{2}+b_0\cos(ky)\right)\right]^2E_{00}+\theta_3\left[\left(-\frac{x^2}{2}+b_0\cos(ky)\right)\right]E_{00}-\theta_4\left[\exp(\alpha|A_z|^2-1\right]E_{00}=0, \tag{29a}$$

$$\frac{2}{k_{\parallel}}\frac{\partial E_{00}}{\partial z}+2k_{\parallel}\theta_1\lambda_e^2\left(\frac{\partial E_{00}}{\partial x}\frac{\partial s}{\partial x}+\frac{\partial E_{00}}{\partial y}\frac{\partial s}{\partial y}\right)+k_{\parallel}\lambda_e^2\theta_1\left(\frac{\partial^2 s}{\partial x^2}+\frac{\partial^2 s}{\partial y^2}\right)E_{00}+E_{00}\exp(k_iz)=0. \tag{29b}$$

Further assuming the Gaussian profile for initial distribution of wave, i.e.,

$$E_{00}^2=\frac{E_{00}'^2}{f_1(z)f_2(z)}\exp\left(\frac{-x^2}{r_{01}^2}-\frac{y^2}{r_{01}^2}+k_iz\right),\qquad s=\beta_1(z)\frac{x^2}{2}+\beta_2(z)\frac{y^2}{2}+\phi(z),. \tag{30}$$

Using Eq. (30) in Eq. (29b), and collecting the terms of x^2 and y^2 we obtain the following equations:

$$\beta_1(z)=\frac{1}{k_{\parallel}^2\theta_1\lambda_e^2}\frac{1}{f_1}\frac{\partial f_1}{\partial z},\quad \beta_2(z)=\frac{1}{k_{\parallel}^2\theta_1\lambda_e^2}\frac{1}{f_2}\frac{\partial f_2}{\partial z}. \tag{31}$$

Here, r_{01} and r_{02} represent the current sheet size of whistler wave in x and y directions, respectively; $\beta_1(z)$ and $\beta_2(z)$ represent the slowly varying function, and f_1 and f_2 represent the beam width parameter of the whistler wave. Now, for obtaining the beam width parameters, use Eq. (30) in Eq. (29a) with paraxial approximation, i.e., $x\ll r_{01}^2f_1, y\ll r_{02}^2f_2$ and then equating coefficients of x^2 and y^2, we obtained the following equation:

$$\frac{\partial^2 f_1}{\partial\xi^2}=R_d^2k_{\parallel}^2\left(\frac{4\theta_1^2\lambda_e^4}{r_{01}^4f_1^4f_2}+\frac{\theta_1\theta_3f_1}{2}-\frac{2\theta_1\theta_4\lambda_e^2\left(\alpha E_{00}'^2\right)}{r_{01}^2f_1^3f_2}\right) \tag{32}$$

and

$$\frac{\partial^2 f_2}{\partial\xi^2}=R_d^2k_{\parallel}^2\left(\frac{4\theta_1^2\lambda_e^4}{r_{02}^4f_2^4f_1}+\frac{\theta_1\theta_3k^2\lambda_e^2f_2}{2}-\frac{\theta_1\theta_4\lambda_e^2\left(\alpha E_{00}'^2\right)}{r_{02}^2f_2^3f_1}\right), \tag{33}$$

where $\xi=z/R_d$ and $R_d=k_{0z}r_{01}^2$. Equations (32) and (33) are solved for the plain wavefront imposing the following initial conditions: $\frac{\partial^2 f_1}{\partial\xi^2}=\frac{\partial^2 f_2}{\partial\xi^2}=0$, $f_1=f_2=1$ at $z=0$. When power of whistler wave $(\alpha E_{00}'^2)$ is zero, then Eqs. (32) and (33) becomes quadratic and gives the transverse scale size of localized structures of whistler wave along x-axis and y-axis, i.e., $r_{01}=7.5\lambda_e$, $r_{02}=5.2\lambda_e$. Furthermore, when we take finite value of power $(\alpha E_{00}'^2=1)$, then the transverse scale size of localized structures of whistler wave along x-axis and y-axis, i.e., $r_{01}=0.8\lambda_e$, $r_{02}=0.63\lambda_e$. Also, by using Ampère's law $(J=-\nabla^2E_{00}')$, we have studied the current sheet size. One can calculate the transverse scale size of current sheet comes out $\frac{r_{01}}{\sqrt{3}}$ and $\frac{r_{02}}{\sqrt{3}}$ along x-axis and y-axis, respectively. Therefore, transverse scale size of current sheet is $4.3\lambda_e$, and $3\lambda_e$ along x-axis and y-axis, respectively. As a result, it shows how the whistler power also influences the scale size of coherent structures and current sheet dimension. Recently, Leonenko *et al.*[72] reported the scale size of superthin electron-scale current sheet in earth's magnetotail. Also, Bruch *et al.*[73] reported the scale size of whistler's localized structures of the order of $(10-20)\lambda_e$ at the magnetic reconnection sites in the magnetopause.

V. SUMMARY AND CONCLUSION

Numerous waves have been observed so far in the reconnection region. The magnetic reconnection zone near the dayside magnetopause exhibits implications of beam-driven whistler mode waves, according to observed data from the MMS.[55] This article deals with the study of beam-driven whistler wave dynamics due to modified background density and with the modified background field of fully developed chain of magnetic islands. Furthermore, these equations are normalized and resolved using the finite difference method and the pseudospectral method. The outcomes of the numerical simulations illustrate that the nonlinear whistler wave propagating in the magnetic islands gets localized and current sheets are formed having the transverse scale size of the order electron inertial length. Although, Stenflo *et al.*[59] also investigated the nonlinear coupling of waves but in the absence of magnetic islands. Also, they have considered the linear growth rate of the system, and their results are relevant to non-thermal magnetic field fluctuations which lead to plasma heating. However, in our model, we studied the nonlinear saturation of waves in the presence of magnetic islands. The outcomes of the numerical simulation are summarized here.

(1) Simulation results reveal the localization of whistler wave. The resolution of the numerical model under consideration is of the order $2.15\lambda_e$, i.e., $(\{\text{system length}\times x_n\}/\text{grid size})$, which is appropriate enough to resolve the structures which are formed. And, in our analysis, localized structures are of the order of $3\lambda_e$. These localized structures of the magnetic field offer a lane for energy cascading from large to small scale spatial size.
(2) Contour plot of flux lines can be utilized to reveal the existence of magnetic islands created by reconnection. We then examine the spatial evolution of current sheet of whistler wave with t in the presence of reconnection induced magnetic islands as well as background density fluctuations. Which illustrates that for the very initial stage, the current is having a smooth or symmetric distribution, as time progresses, we observe that the appropriate X–O structures get deformed, and we eventually reach a wholly chaotic pattern. These chaotic structures may be responsible for the creation of whistler turbulence.
(3) We investigated the ensemble averaged spectra of whistler wave to observe the energy cascading from larger to lower scale due

21 February 2024 06:14:50

Published under an exclusive license by AIP Publishing

to localization process. We just give the scaling of reference lines (red and green scaling lines for −1.2 and −7/3 scaling, respectively) and actual power spectra is given by a solid curve line.

Also, we observed that power of whistler wave also impacts the scale size of coherent structures and current sheet dimension. From the semianalytical model, we have also calculated the transverse scale size of whistler's localized structures and the current sheet size. When the power of whistler wave is zero, then the scale size of current sheet is of the order of $7.5\lambda_e$ and $5.2\lambda_e$ along x-axis and along y-axis, respectively. Furthermore, when the power of whistler wave is finite, then scale size of current sheet is of the order of $4.3\lambda_e$, and $3\lambda_e$ along x-axis and along y-axis, respectively.

The outcomes demonstrate whistler turbulence at magnetic reconnection locations produced by the electron beam as well as localized structures and whistler fluctuations that correspond to observations reported by Zhao *et al.*[55] This causes the heating and acceleration of plasma particles in the magnetopause. Moreover, the nonlinear implementations of coupled waves are already ongoing and could further support to the validity of the model. As a result, our study will enable us to gain a better quantitative understanding of whistler-mode waves produced by energetic electron beams in turbulent circumstances.

ACKNOWLEDGMENTS

Jyoti wishes to acknowledge the financial assistance (Award No. 515500) provided by the University Grant Commission (UGC), Government of India. Also, this work was partially supported by the Department of Science and Technology (DST, Grant No. CRG/2022/000546).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Jyoti: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Methodology (equal); Resources (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Suresh C. Sharma:** Formal analysis (equal); Resources (equal); Supervision (equal); Validation (equal); Visualization (equal). **R.P. Sharma:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Methodology (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX: DISPERSION RELATION OF 3D WHISTLER WAVE

The dynamical equation for a 3D whistler wave with an ambient magnetic field along the z-axis in a magnetized plasma propagating in the x, y, and z planes with a wave vector $\vec{k} = k_x\hat{x} + k_y\hat{y} + k_z\hat{z}$, is derived utilizing two fluid models.

To establish the dynamics of whistler wave, the following key equations are used:

Wave equation:

$$\nabla^2\vec{E} - \nabla(\nabla.\vec{E}) = \frac{4\pi}{c^2}\frac{\partial\vec{J}}{\partial t} + \frac{1}{c^2}\frac{\partial^2\vec{E}}{\partial t^2}. \tag{A1}$$

Equation of motion for electrons:

$$m_e\left[\frac{\partial\vec{v}_e}{\partial t} + (\vec{v}_e.\nabla\vec{v}_e)\right] = -e\vec{E} - \frac{e}{c}(\vec{v}_e \times \vec{B}_0). \tag{A2}$$

Equation of motion for ions:

$$m_i\left[\frac{\partial\vec{v}_i}{\partial t} + (\vec{v}_i.\nabla\vec{v}_i)\right] = e\vec{E} + \frac{e}{c}(\vec{v}_i \times \vec{B}_0), \tag{A3}$$

where m_e m_i denotes the mass of an electron and ion, v_i, v_e denotes the velocity of the ion and electron, B_0 denotes the background magnetic field, and e denotes the charge on the electron, where $\vec{B}_0(= B_{0z}\hat{z} + B_{0y}\hat{y})$, denotes the background magnetic field along z direction and also B_{0y} denotes the magnetic field owing to the presence of magnetic islands (at the reconnection sites).

Velocity components of whistler wave are derived using Eqs. (A2) and (A3),

$$\frac{d\vec{v}_e}{dt} = \frac{\partial\vec{v}_e}{\partial t} + \vec{v}_e.(\nabla\vec{v}_e) = \frac{-e\vec{E}}{m_e} - \frac{e}{m_e c}\left[\vec{v}_e \times \left(B_{0z}\hat{z} + B_{0y}\hat{y}\right)\right], \tag{A4}$$

where $\vec{v}_e = \vec{v}_{e0} + \vec{v}_{e1}$ and $\vec{v}_{e0} = 0 =$ No drift in equilibrium $\vec{B}(= B_{0z}\hat{z} + B_{0y}\hat{y})$, background magnetic field.

We can write the x-component of Eq. (A4) as

$$-i\omega_0\, v_{e1x} = -\frac{eE_x}{m_e} - v_{e1y}\omega_{ce} + \frac{e}{m_e c}v_{e1z}B_{0y}. \tag{A5}$$

Similarly, y-component of Eq. (A4) is

$$-i\omega_0\, v_{e1y} = -\frac{eE_y}{m_e} + v_{e1x}\omega_{ce}. \tag{A6}$$

z-component of Eq. (A4) is

$$-i\omega_0\, v_{e1z} = -\frac{eE_z}{m_e} - \frac{e}{m_e c}v_{e1x}B_{0y}. \tag{A7}$$

From Eq. (A6),

$$v_{e1y} = \frac{eE_y}{i\omega_0 m_e} + \frac{v_{e1x}\omega_{ce}}{i\omega_0}. \tag{A8}$$

Using Eqs. (A8) and (A7) in Eq. (A5), we can obtain the x-component of whistler wave's velocity due to electron.

$$v_{e1x} = -\frac{i\omega_0 e}{m_e\left(\omega_0^2 - \omega_{ce}^2\right)}E_x - \frac{e\omega_{ce}}{m_e\left(\omega_0^2 - \omega_{ce}^2\right)}E_y + \frac{e^2B_{0y}}{m_e^2 c\left(\omega_0^2 - \omega_{ce}^2\right)}E_z. \tag{A9}$$

y-component of whistler wave's velocity due to electron,

21 February 2024 06:14:50

Published under an exclusive license by AIP Publishing

$$v_{e1y} = \frac{e\omega_{ce}}{m_e\left(\omega_0^2 - \omega_{ce}^2\right)}E_x - \frac{i\omega_0 e}{m_e\left(\omega_0^2 - \omega_{ce}^2\right)}E_y + \frac{ie^2\omega_{ce}B_{0y}}{m_e^2 c\omega_0\left(\omega_0^2 - \omega_{ce}^2\right)}E_z. \tag{A10}$$

z-component of whistler wave's velocity due to electron,

$$v_{e1z} = \frac{eE_z}{im_e\omega_0} + \left[\frac{-e^2B_{0y}}{m_e^2 c\left(\omega_0^2 - \omega_{ce}^2\right)}E_x + \frac{ie^2\omega_{ce}B_{0y}}{m_e^2 c\omega_0\left(\omega_0^2 - \omega_{ce}^2\right)}E_y\right]. \tag{A11}$$

Similarly, we can obtain the components of whistler wave's velocity due to ions,

$$v_{i1x} = \frac{i\omega_0 e}{m_i\left(\omega_0^2 - \omega_{ci}^2\right)}E_x - \frac{e\omega_{ci}}{m_i\left(\omega_0^2 - \omega_{ci}^2\right)}E_y - \frac{e^2B_{0y}}{m_i^2 c\left(\omega_0^2 - \omega_{ci}^2\right)}E_z, \tag{A12}$$

$$v_{i1y} = \frac{e\omega_{ci}}{m_i\left(\omega_0^2 - \omega_{ci}^2\right)}E_x + \frac{i\omega_0 e}{m_i\left(\omega_0^2 - \omega_{ci}^2\right)}E_y + \frac{ie^2\omega_{ci}B_{0y}}{m_i^2 c\omega_0\left(\omega_0^2 - \omega_{ci}^2\right)}E_z, \tag{A13}$$

and

$$v_{i1z} = \frac{-eE_z}{im_i\omega_0} + \left[\frac{-e^2B_{0y}}{m_i^2 c\left(\omega_0^2 - \omega_{ci}^2\right)}E_x + \frac{ie^2\omega_{ci}B_{0y}}{m_i^2 c\omega_0\left(\omega_0^2 - \omega_{ci}^2\right)}E_y\right]. \tag{A14}$$

Now, use these velocity components in driving the current density, by using the following relation, $\vec{J} = n_i e\vec{v}_i - n_e e\vec{v}_e$, and then substitute in Eq. (A1):x-component of Eq. (A1),

$$\left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{\left(\omega_0^2 - \omega_{ci}^2\right)\lambda_i^2}\frac{\partial^2}{\partial t^2} + \frac{1}{\left(-\frac{\partial^2}{\partial t^2} - \omega_{ce}^2\right)\lambda_e^2}\frac{\partial^2}{\partial t^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)E_x + \left(-\frac{\partial}{\partial x}\frac{\partial}{\partial y} - \frac{\omega_{ci}}{\left(-\frac{\partial^2}{\partial t^2} - \omega_{ci}^2\right)\lambda_i^2}\frac{\partial}{\partial t} - \frac{\omega_{ce}}{\left(-\frac{\partial^2}{\partial t^2} - \omega_{ce}^2\right)\lambda_e^2}\frac{\partial}{\partial t}\right)E_y + \left(-\frac{\partial}{\partial x}\frac{\partial}{\partial z} + \frac{ie\omega_0 B_{0y}}{m_i c\left(\omega_0^2 - \omega_{ci}^2\right)\lambda_i^2} + \frac{ie\omega_0 B_{0y}}{m_e c\left(\omega_0^2 - \omega_{ce}^2\right)\lambda_e^2}\right)E_z = 0. \tag{A15}$$

y-component Eq. (A1),

$$\left(-\frac{\partial}{\partial y}\frac{\partial}{\partial x} - \frac{\omega_{ci}}{\left(-\frac{\partial^2}{\partial t^2} - \omega_{ci}^2\right)\lambda_i^2}\frac{\partial}{\partial t} + \frac{\omega_{ce}}{\left(-\frac{\partial^2}{\partial t^2} - \omega_{ce}^2\right)\lambda_e^2}\frac{\partial}{\partial t}\right)E_x + \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + \frac{1}{\left(-\frac{\partial^2}{\partial t^2} - \omega_{ci}^2\right)\lambda_i^2}\frac{\partial^2}{\partial t^2} + \frac{1}{\lambda_e^2\left(-\frac{\partial^2}{\partial t^2} - \omega_{ce}^2\right)}\frac{\partial^2}{\partial t^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)E_y \times \left(-\frac{\partial}{\partial y}\frac{\partial}{\partial z} - \frac{e\omega_{ce}B_{oy}}{m_e c\lambda_e^2\left(-\frac{\partial^2}{\partial t^2} - \omega_{ce}^2\right)} + \frac{e\omega_{ci}B_{oy}}{m_i c\lambda_i^2\left(-\frac{\partial^2}{\partial t^2} - \omega_{ci}^2\right)}\right)E_z = 0. \tag{A16}$$

z-component Eq. (A1),

$$\left(-\frac{\partial}{\partial x}\frac{\partial}{\partial z} - \frac{ie\omega_0 B_{0y}}{m_e c\left(-\frac{\partial^2}{\partial t^2} - \omega_{ce}^2\right)} + \frac{ie\omega_0 B_{0y}}{m_i c\left(-\frac{\partial^2}{\partial t^2} - \omega_{ci}^2\right)}\right)E_x + \left(-\frac{\partial}{\partial y}\frac{\partial}{\partial z} - \frac{e\omega_{ce}B_{0y}}{m_e c\left(-\frac{\partial^2}{\partial t^2} - \omega_{ce}^2\right)} + \frac{e\omega_{ci}B_{0y}}{m_i c\left(-\frac{\partial^2}{\partial t^2} - \omega_{ci}^2\right)}\right)E_y \times \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} - \frac{1}{\lambda_e^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)E_z = 0., \tag{A17}$$

where $\omega_{ce}\left(=\frac{eB_0}{m_e c}\right)$ is the electron gyrofrequency of wave, n_0 denotes the background number density, c denotes the velocity of light, and $\lambda_e\left(=\sqrt{\frac{c^2 m_e}{4\pi n_0 e^2}}\right)$ is the collisionless electron skin depth.

Now, compounding Eqs. (A15)–(A17) and applying conditions $k\lambda_i \gg 1$, we get

21 February 2024 06:14:50

Published under an exclusive license by AIP Publishing

$$\left(\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}-\frac{1}{\lambda_e^2\left(-\frac{\partial^2}{\partial t^2}-\omega_{ce}^2\right)}\frac{\partial^2}{\partial t^2}\right)\left[\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}+\frac{1}{\lambda_e^2\left(-\frac{\partial^2}{\partial t^2}-\omega_{ce}^2\right)}\frac{\partial^2}{\partial t^2}\right)\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}-\frac{1}{\lambda_e^2}\right)\right.$$
$$\left.\times\left(\frac{\partial}{\partial y}\frac{\partial}{\partial z}+\frac{e\omega_{ce}B_{0y}}{m_e c\lambda_e^2\left(-\frac{\partial^2}{\partial t^2}-\omega_{ce}^2\right)}\right)^2\right]$$
$$-\left(\frac{\partial}{\partial x}\frac{\partial}{\partial y}+\frac{\omega_{ce}}{\lambda_e^2\left(-\frac{\partial^2}{\partial t^2}-\omega_{ce}^2\right)}\frac{\partial}{\partial t}\right)\left[\left(\frac{\partial}{\partial y}\frac{\partial}{\partial z}+\frac{e\omega_{ce}B_{0y}}{m_e c\lambda_e^2\left(-\frac{\partial^2}{\partial t^2}-\omega_{ce}^2\right)}\right)\left(\frac{\partial}{\partial x}\frac{\partial}{\partial z}+\frac{eB_{0y}}{m_e c\lambda_e^2\left(-\frac{\partial^2}{\partial t^2}-\omega_{ce}^2\right)}\right)\right.$$
$$\left.\times\left(\frac{\partial}{\partial x}\frac{\partial}{\partial y}+\frac{\omega_{ce}}{\lambda_e^2\left(-\frac{\partial^2}{\partial t^2}-\omega_{ce}^2\right)}\frac{\partial}{\partial t}\right)\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}-\frac{1}{\lambda_e^2}\right)\right]$$
$$+\left(\frac{\partial}{\partial x}\frac{\partial}{\partial z}+\frac{eB_{0y}}{m_e c\lambda_e^2\left(-\frac{\partial^2}{\partial t^2}-\omega_{ce}^2\right)}\frac{\partial}{\partial t}\right)\left[\left(\frac{\partial}{\partial x}\frac{\partial}{\partial y}+\frac{\omega_{ce}}{\lambda_e^2\left(-\frac{\partial^2}{\partial t^2}-\omega_{ce}^2\right)}\frac{\partial}{\partial t}\right)\left(\frac{\partial}{\partial y}\frac{\partial}{\partial z}+\frac{e\omega_{ce}B_{0y}}{m_e c\lambda_e^2\left(-\frac{\partial^2}{\partial t^2}-\omega_{ce}^2\right)}\right)\right.$$
$$\left.\times\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}+\frac{1}{\lambda_e^2\left(-\frac{\partial^2}{\partial t^2}-\omega_{ce}^2\right)}\frac{\partial^2}{\partial t^2}\right)\left(\frac{\partial}{\partial x}\frac{\partial}{\partial z}+\frac{eB_{0y}}{m_e c\lambda_e^2\left(-\frac{\partial^2}{\partial t^2}-\omega_{ce}^2\right)}\frac{\partial}{\partial t}\right)\right]E_z=0. \tag{A18}$$

After solving Eq. (A18), we obtain the dynamics of 3D whistler wave in the presence of magnetic island in terms of electric field under the condition $-\frac{\partial^2}{\partial t^2}>\omega_{ce}\omega_{ci}$. Now, we can write the dynamical equation of 3D whistler in terms of electric field.

$$\frac{\partial^2 E_z}{\partial t^2}+\frac{\partial^6 E_z}{\partial t^2\partial x^4}\lambda_e^4+\frac{\partial^6 E_z}{\partial t^2\partial y^4}\lambda_e^4+\frac{\partial^6 E_z}{\partial t^2\partial z^4}\lambda_e^4+2\lambda_e^4\frac{\partial^4 E_z}{\partial t^2\partial z^2}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)-2\lambda_e^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\frac{\partial^2 E_z}{\partial t^2}$$
$$+\lambda_i^2 v_A^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\frac{\partial^2 E_z}{\partial z^2}+\lambda_i^2 v_A^2\frac{\partial^4 E_z}{\partial z^4}+\frac{\lambda_i^2 v_A^2 B_{0y}^2}{B_{0z}^2}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\frac{\partial^2 E_z}{\partial y^2}+\frac{\lambda_i^2 v_A^2 B_{0y}^2}{B_{0z}^2}\frac{\partial^4 E_z}{\partial y^2\partial z^2}$$
$$+\frac{2\lambda_i^2 v_A^2 B_{0y}}{B_{0z}}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\frac{\partial^2 E_z}{\partial y\partial z}+\frac{2\lambda_i^2 v_A^2 B_{0y}}{B_{0z}}\frac{\partial^4 E_z}{\partial y\partial z^3}=0. \tag{A19}$$

Also, we can write above equation in the following form, which is the dispersion relation in the presence of magnetic islands:

$$\left(\omega_0^2(1+k^2\lambda_e^2)^2-\frac{k^2\lambda_i^2(k_z B_{0z}+k_y B_{0y})^2}{4\pi n_0 m_i}\right)E_z=0. \tag{A20}$$

Here, $B_0(=B_{0z}\hat{z}+B_{0y}\hat{y})$ denotes the background magnetic field along z direction and also B_{0y} denotes the magnetic field owing to the presence of magnetic islands (at the reconnection sites), and in the absence of magnetic islands, one can get the usual dispersion relation of whistler wave as follows:

$$\left[\omega_0^2(1+k^2\lambda_e^2)^2-k^2k_z^2\lambda_i^2v_A^2\right]E_z=0. \tag{A21}$$

Moreover, it is emphasized that the equation of motion is then used to examine beam-driven evolution.

$$m\frac{dv}{dt}=-eE-mv\gamma \quad \rightarrow \frac{dv}{dt}=-\frac{eE}{m}-v\gamma,$$

$$(-i\omega+\gamma)=-\frac{eE}{mv} \quad \rightarrow \frac{d}{dt}=\left(\frac{d}{dt}+\gamma\right).$$

$\frac{d^2}{dt^2}=\left(\frac{d^2}{dt^2}+2\gamma\frac{d}{dt}\right)$, neglecting higher order terms.

The dispersion relation of the whistler wave under the conditions,

$$k_{0x}=k_{0y}=0, \quad k_0=k_{0z} \quad \text{and} \quad k_0\lambda_e\gg 1, \quad \omega_{ce}>\omega_0.$$

Equation (A21) modifies to

$$\omega_0^2=k_{0z}^4\lambda_i^2v_A^2. \tag{A22}$$

It should be noted that this formulation [Eq. (A22)] is identical to the dispersion relation of a circularly polarized wave.[58] (The widely recognized dispersion function of a circularly polarized whistler wave[58] is given as

$$k_0^2=\frac{\omega_0^2}{c^2}\left(1-\frac{\omega_{pe}^2}{\omega_0(\omega_0-\omega_{ce})}\right).$$

21 February 2024 06:14:50

Published under an exclusive license by AIP Publishing

Now, after neglecting the displacement current and taking the conditions $k_{0x} = k_{0y} = 0, k_0 = k_{0z}$, $k_0\lambda_e \gg 1$, and $\omega_{ce} > \omega_0$ into account, it reduces $\omega_0 = k_{0z}^2\lambda_e^2\omega_{ce}$. Furthermore, using $\lambda_i v_A = \lambda_e^2\omega_{ce}$ and hence the similarity.

REFERENCES

[1]J. Birn and E. R. Priest, *Reconnection of Magnetic Fields: Magnetohydrodynamics and Collisionless Theory and Observations* (Cambridge University Press, 2007).
[2]M. Yamada, J. Yoo, J. Jara-Almonte, H. Ji, R. M. Kulsrud, and C. E. Myers, "Conversion of magnetic energy in the magnetic reconnection layer of a laboratory plasma," Nat. Commun. **5**, 4774 (2014).
[3]G. B. Field, "Magnetic Effects: Cosmical Magnetic Fields. Their Origin and Their Activity E. N. Parker. Clarendon (Oxford University Press), New York, 1979. xviii, 842 pp., illus. $95. International Series of Monographs on Physics," Science **207**(4432), 753 (1980).
[4]R. M. Kulsrud, "Magnetic reconnection in a magnetohydrodynamic plasma," Phys. Plasmas **5**, 1599–1606 (1998).
[5]T. Li, E. Priest, and R. Guo, "Three-dimensional magnetic reconnection in astrophysical plasmas," Proc. R. Soc. A **477**(2249), 20200949 (2021).
[6]D. Biskamp, E. Schwarz, A. Zeiler, A. Celani, and J. F. Drake, "Electron magnetohydrodynamic turbulence," Phys Plasmas **6**(3), 751–758 (1999).
[7]W. Gonzalez and E. Parker, *Magnetic Reconnection Concepts and Applications* (Springer, 1994).
[8]M. Yamada, R. Kulsrud, and H. Ji, "Magnetic reconnection," Rev. Mod. Phys. **82**(1), 603 (2010).
[9]R. E. Ergun, K. A. Goodrich, F. D. Wilder, N. Ahmadi, J. C. Holmes, S. Eriksson, J. E. Stawarz, R. Nakamura, K. J. Genestreti, M. Hesse, J. L. Burch, R. B. Torbert, T. D. Phan, S. J. Schwartz, J. P. Eastwood, R. J. Strangeway, O. Le Contel, C. T. Russell, M. R. Argall, P. A. Lindqvist, L. J. Chen, P. A. Cassak, B. L. Giles, J. C. Dorelli, D. Gershman, T. W. Leonard, B. Lavraud, A. Retino, W. Matthaeus, and A. Vaivads, "Magnetic reconnection, turbulence, and particle acceleration: Observations in the Earth's magnetotail," Geophys. Res. Lett. **45**(8), 3338–3347, https://doi.org/10.1002/2018GL076993 (2018).
[10]M. Hoshino and Y. Lyubarsky, "Relativistic reconnection and particle acceleration," Space Sci. Rev. **173**, 521–533 (2012).
[11]G. R. Werner, D. A. Uzdensky, M. C. Begelman, B. Cerutti, and K. Nalewajko, "Non-thermal particle acceleration in collisionless relativistic electron–proton reconnection," Mon. Not. R. Astron. Soc. **473**(4), 4840–4861 (2018).
[12]S. Lu, V. Angelopoulos, A. V. Artemyev, P. L. Pritchett, J. Liu, A. Runov, A. Tenerani, C. Shi, and M. Velli, "Turbulence and particle acceleration in collisionless magnetic reconnection: Effects of temperature inhomogeneity across pre-reconnection current sheet," Astrophys. J. **878**(2), 109 (2019).
[13]G. R. Werner and D. A. Uzdensky, "Nonthermal particle acceleration in 3D relativistic magnetic reconnection in pair plasma," Astrophys. J. **843**(2), L27 (2017).
[14]R. E. Ergun, N. Ahmadi, L. Kromyda, S. J. Schwartz, A. Chasapis, S. Hoilijoki, F. D. Wilder, J. E. Stawarz, K. A. Goodrich, D. L. Turner, I. J. Cohen, S. T. Bingham, J. C. Holmes, R. Nakamura, F. Pucci, R. B. Torbert, J. L. Burch, P.-A. Lindqvist, R. J. Strangeway, O. Le Contel, and B. L. Giles, "Observations of particle acceleration in magnetic reconnection–driven turbulence," Astrophys. J. **898**(2), 154 (2020).
[15]D. Kagan, E. Nakar, and T. Piran, "Physics of the saturation of particle acceleration in relativistic magnetic reconnection," Mon. Not. R. Astron. Soc. **476**(3), 3902–3912 (2018).
[16]D. Biskamp, "Magnetic reconnection in plasmas," Astrophys. Space Sci. **242**, 165–207 (1996).
[17]T. K. M. Nakamura, W. L. Teh, S. Zenitani, T. Umeda, M. Oka, H. Hasegawa, A. M. Veronig, and R. Nakamura, "Spatial and time scaling of coalescing multiple magnetic islands," Phys. Plasmas **30**(2), 022902 (2023).
[18]J. L. Burch, T. E. Moore, R. B. Torbert, and B. L. Giles, "Magnetospheric multiscale overview and science objectives," Space Sci. Rev. **199**, 5–21 (2016).
[19]H. Karimabadi and A. Lazarian, "Magnetic reconnection in the presence of externally driven and self-generated turbulence," Phys. Plasmas **20**(11), 112102 (2013).
[20]H. Karimabadi, V. Roytershteyn, W. Daughton, and Y. H. Liu, "Recent evolution in the theory of magnetic reconnection and its connection with turbulence," Space Sci. Rev. **178**(2–4), 307–323 (2013).
[21]H. Karimabadi, V. Roytershteyn, H. X. Vu, Y. A. Omelchenko, J. Scudder, W. Daughton, A. Dimmock, K. Nykyri, M. Wan, D. Sibeck, M. Tatineni, A. Majumdar, B. Loring, and B. Geveci, "The link between shocks, turbulence, and magnetic reconnection in collisionless plasmas," Phys. Plasmas **21**(6), 062308 (2014).
[22]A. Lazarian, G. L. Eyink, and E. T. Vishniac, "Relation of astrophysical turbulence and magnetic reconnection," Phys. Plasmas **19**(1), 012105 (2012).
[23]Z. Vörös, E. Yordanova, A. Varsani, K. J. Genestreti, Y. V. Khotyaintsev, W. Li, D. B. Graham, C. Norgren, R. Nakamura, Y. Narita, F. Plaschke, W. Magnes, W. Baumjohann, D. Fischer, A. Vaivads, E. Eriksson, P. A. Lindqvist, G. Marklund, R. E. Ergun, M. Leitner, M. P. Leubner, R. J. Strangeway, O. Le Contel, C. Pollock, B. J. Giles, R. B. Torbert, J. L. Burch, L. A. Avanov, J. C. Dorelli, D. J. Gershman, W. R. Paterson, B. Lavraud, and Y. Saito, "MMS observation of magnetic reconnection in the turbulent magnetosheath," J. Geophys. Res. **122**(11), 442, https://doi.org/10.1002/2017JA024535 (2017).
[24]A. Retinò, D. Sundkvist, A. Vaivads, F. Mozer, M. André, and C. J. Owen, "*In situ* evidence of magnetic reconnection in turbulent plasma," Nat. Phys. **3**(4), 235–238 (2007).
[25]G. Kowal, D. A. Falceta-Gonçalves, A. Lazarian, and E. T. Vishniac, "Statistics of reconnection-driven turbulence," Astrophys. J. **838**(2), 91 (2017).
[26]W. H. Matthaeus and S. L. Lamkin, "Turbulent magnetic reconnection," Phys. Fluids **29**(8), 2513–2534 (1986).
[27]J. P. Eastwood, T. D. Phan, S. D. Bale, and A. Tjulin, "Observations of turbulence generated by magnetic reconnection," Phys. Rev. Lett. **102**(3), 035001 (2009).
[28]A. Mallet and A. A. Schekochihin, "A statistical model of three-dimensional anisotropy and intermittency in strong Alfvénic turbulence," Mon. Not. R. Astron. Soc. **466**(4), 3918–3927 (2017).
[29]S. Boldyrev, "Spectrum of magnetohydrodynamic turbulence," Phys. Rev. Lett. **96**(11), 115002 (2006).
[30]A. Mallet, A. A. Schekochihin, and B. D. G. Chandran, "Disruption of sheet-like structures in Alfvénic turbulence by magnetic reconnection," Mon. Not. R. Astron. Soc. **468**(4), 4862–4871 (2017).
[31]P. Goldreich and S. Sridhar, "Toward a theory of interstellar turbulence. II. Strong Alfvenic turbulence," Astrophys. J. **468**, 763 (1995).
[32]S. Servidio, W. H. Matthaeus, M. A. Shay, P. A. Cassak, and P. Dmitruk, "Magnetic reconnection in two-dimensional magnetohydrodynamic turbulence," Phys. Rev. Lett. **102**(11), 115003 (2009).
[33]N. F. Loureiro and S. Boldyrev, "Role of magnetic reconnection in magnetohydrodynamic turbulence," Phys. Rev. Lett. **118**, 245101 (2017).
[34]L. Franci, S. S. Cerri, F. Califano, S. Landi, E. Papini, A. Verdini, L. Matteini, F. Jenko, and P. Hellinger, "Magnetic reconnection as a driver for a sub-ion-scale cascade in plasma turbulence," Astrophys. J. **850**(1), L16 (2017).
[35]S. Adhikari, T. N. Parashar, M. A. Shay, W. H. Matthaeus, P. S. Pyakurel, S. Fordin, J. E. Stawarz, and J. P. Eastwood, "Energy transfer in reconnection and turbulence," Phys. Rev. E **104**(6), 065206 (2021).
[36]S. Adhikari, M. A. Shay, T. N. Parashar, W. H. Matthaeus, P. S. Pyakurel, J. E. Stawarz, and J. P. Eastwood, "Effect of a guide field on the turbulence like properties of magnetic reconnection," Phys. Plasmas **30**, 1–15 (2023).
[37]X. H. Wei, J. B. Cao, G. C. Zhou, O. Santolík, H. Rème, I. Dandouras, N. Cornilleau-Wehrlin, E. Lucek, C. M. Carr, and A. Fazakerley, "Cluster observations of waves in the whistler frequency range associated with magnetic reconnection in the Earth's magnetotail," J. Geophys. Res. **112**(10), A10, https://doi.org/10.1029/2006JA011771 (2007).
[38]X. Yu, Q. Lu, R. Wang, X. Gao, L. Sang, and S. Wang, "Simultaneous observation of whistler waves and electron cyclotron harmonic waves in the separatrix region of magnetopause reconnection," J. Geophys. Res. **126**(10), e2021JA029609, https://doi.org/10.1029/2021JA029609 (2021).
[39]S. Y. Huang, H. S. Fu, Z. G. Yuan, A. Vaivads, Y. V. Khotyaintsev, A. Retino, M. Zhou, D. B. Graham, K. Fujimoto, F. Sahraoui, X. H. Deng, B. Ni, Y. Pang, S. Fu, D. D. Wang, and X. Zhou, "Two types of whistler waves in the hall reconnection region," J. Geophys. Res. **121**(7), 6639–6646, https://doi.org/10.1002/2016JA022650 (2016).

21 February 2024 06:14:50

Published under an exclusive license by AIP Publishing

[40]F. D. Wilder, R. E. Ergun, K. A. Goodrich, M. V. Goldman, D. L. Newman, D. M. Malaspina, A. N. Jaynes, S. J. Schwartz, K. J. Trattner, J. L. Burch, M. R. Argall, R. B. Torbert, P. A. Lindqvist, G. Marklund, O. Le Contel, L. Mirioni, Y. V. Khotyaintsev, R. J. Strangeway, C. T. Russell, C. J. Pollock, B. L. Giles, F. Plaschke, W. Magnes, S. Eriksson, J. E. Stawarz, A. P. Sturner, and J. C. Holmes, "Observations of whistler mode waves with nonlinear parallel electric fields near the dayside magnetic reconnection separatrix by the Magnetospheric Multiscale mission," Geophys Res. Lett. **43**(12), 5909–5917 (2016).
[41]S. Y. Huang, Z. G. Yuan, H. S. Fu, A. Vaivads, F. Sahraoui, Y. V Khotyaintsev, A. Retino, M. Zhou, D. Graham, K. Fujimoto, X. H. Deng, B. B. Ni, Y. Pang, S. Fu, and D. D. Wang, in 2nd URSI Atlantic Radio Science Meeting (AT-RASC 2018), 2018.
[42]D. Cao, H. S. Fu, J. B. Cao, T. Y. Wang, D. B. Graham, Z. Z. Chen, F. Z. Peng, S. Y. Huang, Y. V. Khotyaintsev, M. André, C. T. Russell, B. L. Giles, P. A. Lindqvist, R. B. Torbert, R. E. Ergun, O. Le Contel, and J. L. Burch, "MMS observations of whistler waves in electron diffusion region," Geophys. Res. Lett. **44**(9), 3954–3962, https://doi.org/10.1002/2017GL072703 (2017).
[43]X. H. Deng and H. Matsumoto, "Rapid magnetic reconnection in the Earth's magnetosphere mediated by whistler waves," Nature **410**(6828), 557–560 (2001).
[44]H. Viberg, Y. V. Khotyaintsev, A. Vaivads, M. André, and J. S. Pickett, "Mapping HF waves in the reconnection diffusion region," Geophys. Res. Lett. **40**(6), 1032–1037, https://doi.org/10.1002/grl.50227 (2013).
[45]C. C. Chaston, T. D. Phan, J. W. Bonnell, F. S. Mozer, M. Acuña, M. L. Goldstein, A. Balogh, M. Andre, H. Reme, and A. Fazakerley, "Drift-kinetic Alfvén waves observed near a reconnection X line in the Earth's magnetopause," Phys. Rev. Lett. **95**(6), 065002 (2005).
[46]C. Chaston, J. Bonnell, J. P. McFadden, C. W. Carlson, C. Cully, O. Le Contel, A. Roux, H. U. Auster, K. H. Glassmeier, V. Angelopoulos, and C. T. Russell, "Turbulent heating and cross-field transport near the magnetopause from THEMIS," Geophys. Res. Lett. **35**(17), L17S08, https://doi.org/10.1029/2008GL033601 (2008).
[47]W. M. Farrell, M. D. Desch, K. W. Ogilvie, M. L. Kaiser, and K. Goetz, "The role of upper hybrid waves in magnetic reconnection," Geophys. Res. Lett. **30**(24), 2259, https://doi.org/10.1029/2003GL017549 (2003).
[48]D. B. Graham, A. Vaivads, Y. V. Khotyaintsev, M. André, O. L. Contel, D. M. Malaspina, P. A. Lindqvist, F. D. Wilder, R. E. Ergun, D. J. Gershman, B. L. Giles, W. Magnes, C. T. Russell, J. L. Burch, and R. B. Torbert, "Large-amplitude high-frequency waves at Earth's magnetopause," J. Geophys. Res. **123**(4), 2630–2657, https://doi.org/10.1002/2017JA025034 (2018).
[49]D. B. Graham, Y. V. Khotyaintsev, C. Norgren, A. Vaivads, M. André, J. F. Drake, J. Egedal, M. Zhou, O. L. Contel, J. M. Webster, B. Lavraud, I. Kacem, V. Génot, C. Jacquey, A. C. Rager, D. J. Gershman, J. L. Burch, and R. E. Ergun, "Universality of lower hybrid waves at Earth's magnetopause," J. Geophys. Res. **124**(11), 8727–8760, https://doi.org/10.1029/2019JA027155 (2019).
[50]M. Zhou, X. H. Deng, S. Y. Li, Y. Pang, A. Vaivads, H. Réme, E. Lucek, S. Fu, X. Lin, Z. G. Yuan, and J. F. Wang, "Observation of waves near lower hybrid frequency in the reconnection region with thin current sheet," J. Geophys. Res. **114**(A2), A02216, https://doi.org/10.1029/2008JA013427 (2009).
[51]X. H. Deng, R. X. Tang, H. Matsumoto, J. S. Pickett, A. N. Fazakerley, H. Kojima, W. Baumjohann, A. Coates, R. Nakamura, D. A. Gurnett, and Z. X. Liu, "Observations of electrostatic solitary waves associated with reconnection by Geotail and Cluster," Adv. Space Res. **37**(7), 1373–1381 (2006).
[52]H. Matsumoto, X. H. Deng, H. Kojima, and R. R. Anderson, "Observation of electrostatic solitary waves associated with reconnection on the dayside magnetopause boundary," Geophys. Res. Lett. **30**(6), 1326, https://doi.org/10.1029/2002GL016319 (2003).
[53]C. M. Liu, A. Vaivads, D. B. Graham, Y. V. Khotyaintsev, H. S. Fu, A. Johlander, M. André, and B. L. Giles, "Ion-beam-driven intense electrostatic solitary waves in reconnection jet," Geophys. Res. Lett. **46**(22), 12702–127210, https://doi.org/10.1029/2019GL085419 (2019).
[54]C. Krafft, P. Thévenet, G. Matthieussent, B. Lundin, G. Belmont, B. Lembège, J. Solomon, J. Lavergnat, and T. Lehner, "Whistler wave emission by a modulated electron beam," Phys. Rev. Lett. **72**(5), 649–652 (1994).
[55]S. Q. Zhao, C. J. Xiao, T. Z. Liu, H. Chen, H. Zhang, M. J. Shi, S. Teng, H. S. Zhang, X. G. Wang, Z. Y. Pu, and M. Z. Liu, "Observations of the beam-driven whistler mode waves in the magnetic reconnection region at the dayside magnetopause," J. Geophys. Res. **126**(2), e2020JA028525, https://doi.org/10.1029/2020JA028525 (2021).
[56]A. Vaivads, Y. Khotyaintsev, M. André, and R. A. Treumann, *Plasma Waves Near Reconnection Sites*, Lecture Notes in Physics (LNP, 2006), Vol. 687.
[57]W. Jing, Y. Jiang, R. Aojun, C. Zuzheng, X. Xiaojun, C. Jun, and C. Jinbin, "In situ observations of whistler-mode waves in magnetic reconnection at Mars," Astrophys. J **944**(1), 85 (2023).
[58]S. S. Cerri and F. Califano, "Reconnection and small-scale fields in 2D-3V hybrid-kinetic driven turbulence simulations," New J. Phys. **19**(2), 025007 (2017).
[59]L. Stenflo, M. Y. Yu, and P. K. Shukla, "Electromagnetic modulations of electron whistlers on plasmas," J. Plasma Phys. **36**(3), 447–452 (1986).
[60]Jyoti, S. C. Sharma, N. Pathak, and R. P. Sharma, "Beam-driven whistler mode nonlinear saturation and turbulence in the magnetopause," Phys. Plasmas **29**(9), 092104 (2022).
[61]Jyoti, S. C. Sharma, and R. P. Sharma, "Localization and turbulence of beam-driven whistler wave with magnetosonic wave in magnetopause," Phys. Plasmas **30**(2), 022904 (2023).
[62]R. Fitzpatrick and F. L. Waelbroeck, "Two-fluid magnetic island dynamics in slab geometry. II. Islands interacting with resistive walls or resonant magnetic perturbations," Phys. Plasmas **12**(2), 022308 (2005).
[63]N. Pathak, N. Yadav, S. Sharma, P. Sharma, and R. P. Sharma, "Localization of whistler wave and turbulent spectra in the magnetotail region," J. Geophys. Res. **122**(2), 1751–1762, https://doi.org/10.1002/2016JA023640 (2017).
[64]D. Biskamp and H. Welter, "Dynamics of decaying two-dimensional magneto-hydrodynamic turbulence," Phys. Fluids B **1**(10), 1964–1979 (1989).
[65]S. Boldyrev and N. F. Loureiro, "Role of reconnection in inertial kinetic-Alfvén turbulence," Phys. Rev. Res. **1**(1), 012006 (2019).
[66]S. Ichimaru, "Basic principles of plasma physics: A statistical approach. By S. Ichimaru. Benjamin frontiers in physics," J. Plasma Phys. **13**(3), 571–572 (1975).
[67]M. Colunga, J. F. Luciani, and P. Mora, "Electron acceleration in localized plasma waves," Phys. Fluids **29**(10), 3407–3414 (1986).
[68]V. N. Tsytovich, "Radiative-resonant collective wave-particle interactions," Phys. Rep. **178**(5–6), 261–387 (1989).
[69]R. P. Sharma, H. D. Singh, and M. Malik, "Alfvén wave filamentation and particle acceleration in solar wind and magnetosphere," J. Geophys. Res. **111**(12), A12108, https://doi.org/10.1029/2006JA011759 (2006).
[70]N. H. Bian and P. K. Browning, "Particle acceleration in a model of a turbulent reconnecting plasma: A fractional diffusion approach," Astrophys. J **687**(2), L111 (2008).
[71]W. Rozmus, R. P. Sharma, J. C. Samson, and W. Tighe, "Nonlinear evolution of stimulated Raman scattering in homogeneous plasmas," Phys. Fluids **30**(7), 2181–2193 (1987).
[72]M. V. Leonenko, E. E. Grigorenko, and L. M. Zelenyi, "Spatial scales of super thin current sheets with MMS observations in the Earth's magnetotail," Geomagn. Aeron. **61**(5), 688–695, https://doi.org/10.1134/S0016793221050091 (2021).
[73]J. L. Burch, R. E. Ergun, P. A. Cassak, J. M. Webster, R. B. Torbert, B. L. Giles, J. C. Dorelli, A. C. Rager, K. J. Hwang, T. D. Phan, K. J. Genestreti, R. C. Allen, L. J. Chen, S. Wang, D. Gershman, O. Le Contel, C. T. Russell, R. J. Strangeway, F. D. Wilder, D. B. Graham, M. Hesse, J. F. Drake, M. Swisdak, L. M. Price, M. A. Shay, P. A. Lindqvist, C. J. Pollock, R. E. Denton, and D. L. Newman, "Localized oscillatory energy conversion in magnetopause reconnection," Geophys. Res. Lett. **45**(3), 1237–1245, https://doi.org/10.1002/2017GL076809 (2018).

21 February 2024 06:14:50

Published under an exclusive license by AIP Publishing

Posted on 30 Jan 2024 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.170663531.15269883/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.

Low-Energy Decentralized Context-aware Access Control in Internet of Things

Surendra Tyagi[1], Devesh Jinwala[2], and Subhasis Bhattacharjee[1]

[1]IIT Jammu
[2]SVNIT

January 30, 2024

Abstract

Access control in the Internet of Things plays an important role in keeping resources and services secure. As IoT is expanding, centralized access control will not be able to provide service to the entire network in a timely manner. Along with increasing communication overhead, Centralized access control is not suitable for resource-constrained devices. So, in this paper, we have introduced a Decentralized Context-Aware Access Control (DCAAC) mechanism, which reduces communication overhead, provides fast response, and increases battery life by 41.67 \%. The proposed decentralized access control eliminates all intermediary attacks such as Eavesdropping, snooping, interception, and modification attack, which are usually present in a centralized system.

Received <day> <Month>, <year>; | Revised <day> <Month>, <year>; | Accepted <day> <Month>, <year>

DOI: xxx/xxxx

ARTICLE TYPE

Low-Energy Decentralized Context-aware Access Control in Internet of Things

Surendra Tyagi[1] | Devesh C. Jinwala[2] | Subhasis Bhattacharjee[1]

[1]Department of Computer Science & Engineering, Indian Institute of Technology, Jammu, Jammu & Kashmir, India
[2]Computer Science, National Institute of Technology, Surat, Gujarat, India

Correspondence
*Surendra Tyagi, IIT Jammu. Email: 2018rcs0012@iitjammu.ac.in

Present Address
Indian Institute of Technology, Jammu

Abstract

Access control in the Internet of Things plays an important role in keeping resources and services secure. As IoT is expanding, centralized access control will not be able to provide service to the entire network in a timely manner. Along with increasing communication overhead, Centralized access control is not suitable for resource-constrained devices. So, in this paper, we have introduced a Decentralized Context-Aware Access Control (DCAAC) mechanism, which reduces communication overhead, provides fast response, and increases battery life by 41.67 %. The proposed decentralized access control eliminates all intermediary attacks such as Eavesdropping, snooping, interception, and modification attack, which are usually present in a centralized system.

KEYWORDS:
Internet of Things, IoT, context-based, Access Control, ontology

1 | INTRODUCTION

The word "Internet of things" (IoT) was used for the first time in 1999 by Kevin Ashton[1] . The IoT is a network of devices that have a unique individual identity. All types of devices from low computing devices (sensors, actuators, etc.) to high computing devices (Servers, laptops, etc.) are part of IoT. They can communicate with each other and perform operations without human intervention[2].

IoT has taken a revolution in the field of agriculture, transportation, healthcare, industrial automation, and emergency services during floods, earthquakes, accidents etc.[3]. These things require a new type of access control techniques that can understand situations[4] and relations between entities.

In IoT, devices (intentionally or maliciously) may try to steal information from other devices. It increases the risk of security and privacy breaches and attacks possibility. the main challenges, in IoT security, are access control, Authentication, Privacy, Information Storage and Management, Privacy and System Configuration[5].

Authorization plays important role in information and resource protection. Access Control Mechanism (ACM) is defined as[6] "The logical component that serves to receive the access request from the subject, to decide, and to enforce the access decision". Traditional Access Control such as Discretionary Access Control (DAC), Mandatory Access Control (MAC), Role-Based Access Control (RBAC), and Attribute Access Control (ABAC) are being used in IT systems. In IoT, these techniques are unable to fulfill all requirements such as scalability, management, efficiency, Heterogeneity, Causality, and Lightweight[7].

(I). Scalability: IoT networks are increasing day by day. Some of the devices are connected for a short duration and some of them are for a long duration or permanently. All devices are generating data. When the volume of information increases, the number of users also increases, it becomes a tedious job to handle the whole information and users[8]. Traditional

access control policies are not sufficient for access control. Access Control List (ACL) and RBAC are unable to handle this network properly. Because to maintain a list of so many devices in one place is a very tedious job. If more entities request services, instantly providing authorization to all simultaneously is also a critical job. This is an overloading situation.

(II). Heterogeneity: There are different types of devices with different platforms and technologies are being connected. Every device has its policy for authorization. Traditional access control methods both RBAC and ABAC are inflexible to provide efficient and effective intra-domain authorization services.

(III). Causality: Generally, RBAC and ABAC both work on the permanent network. Some of the devices are connected for a short duration in IoT[9].so dynamic access control techniques are required to handle this network.

(IV). Lightweight: IoT devices are resource constraints. Which don't have heavy computation power and large memory for storage. They work in a lossy communication environment. These devices are connected with low power. It should be minimum to minimum communication to other devices battery can be saved. It requires an access control mechanism that can save the overhead of communication networks[7] .

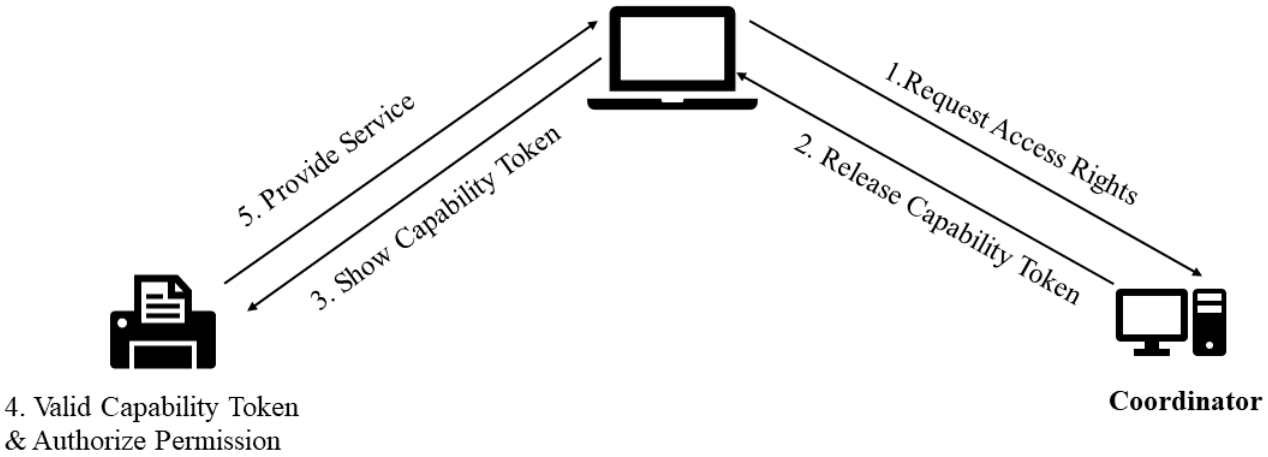


FIGURE 1 System Architecture of FedCAC

Xu et al.[7] suggested Federated Capability-based Access Control (FedCAC) which has capability management, delegation authority, and access right validation. Authorization is being provided by a third party(coordinator), which is a token-based approach. It also requires at least 5 communications between the requester and coordinator, the requester and service provider. Therefore It has also communication overhead.figure 1 shows this communication for Authorization

Mehta et al[10] has proposed A decentralized context-aware access control model for the IoT. This model contains context reasoning, context updation, and context resolver at every device. It has one central device to get any context value that is not available at the local device. Context resolver resolves any conflict related to context. The system administrator defines an ontology for every device. If any local device does not have a context value, the central context manager can provide these values to that device. If the central context manager does not have a requested context value, it sends a request to another device to find the context value. This device sends context value to the central context manager and the central context manager sends this value to the local device which has requested it. Here device is a service provider.

There are some limitations of this model. Local conditions are not taken into account, while local conditions affect security the most. Administrators should formulate access control policies according to local conditions. When the current state of a device is once used for decision-making, then if there is a change in the status of the device, the access to the other device from its changed state should be continued or stopped, any such arrangement is Not there. For example, if door status information is not available at the local device, so it requests a central device and if the central device does not have this value, it requests it to another device. This device sends the door's present status to the central device and the central device sends it to the requested local device. During this transmission someone changes door status, there is no arrangement of decision revoking.

Sofia et al.[11] suggested (Privacy via Anomaly detection System) PALS where smart home is categorized in public room and private room. This is a centralized system. In this, access control is provided based on the private and public rooms without considering family relation context.

To fulfill all requirements, a local authorization device is required, which reduces communication overhead and provides access control based on context. We proposed decentralized access control based on family relation. the main contribution of this paper,

i. We proposed Decentralized Context-Aware Access Control(DCAAC), which is based on family relations or social context for the smart home.

ii. Administrators formulate access control policies according to local conditions.

iii. Context is represented with the help of ontology[12].

iv. Local device gives a quick response. When users are not coming, it will go for deep sleeping mode. It saves energy and increases battery life.

v. In this paper access control is more granular location such as private and public room in smart home.

In our model, each device is uploaded with the access control policies to be executed by that device. The device then executes the policies and grants/rejects the requests based on the satisfaction or otherwise of the access control policy. Since the device is not required to communicate with the central access control resolution module for granting/rejecting access control requests the communication costs in the proposed approach reduces by an order of 41.69 %.

The rest of this paper is organized as follows. Section 2 is the study of related work about context-aware access control. Section 3 gives information about the proposed model. Section 4 is showing Security Analysis in Decentralized System. Section 5 is for Security Against Attacks. Section 6 is for the Experimental setup. Which gives information about the platforms of the experiment. Section 7 is for Energy consumption calculation theory. Section 8 is showing Results and compares them. Section 9 is discussing the given model's relevance. In the last section, there is conclusion.

2 | RELATED WORK

To protect resources from unauthorized users, it requires access control. There are different types of access control Mechanisms. discretionary access control, mandatory access control, role-based access control, attribute-based access control, capability-based Access Control(CapBaC), and context-aware access control.

Discretionary access control restricts[13] access to objects based on the identity of a user or a group to which the user belongs. following are basic building blocks:1. Access Control List 2. DAC policy. Access Control List: It is a list of operations that are to be performed by the subject on the object. In DAC owner objects decide who can access its object-based on policies of access control.

Mandatory Access Control: In this[14], reaching the user is restricted based on the identity and sensitivity of the information that is contained in that object.

Role-based Access Control: a fundamental access control Mechanism in which a user is permitted to access resources based on a role[15]. Following are basic building blocks.1. user 2. role 3. permission 4. assignment policy of user 5. assignment policy of role.

Attribute-based Access Control: When access is provided based on attributes of object and subject. It is called attribute-based access control[6]. There are three building blocks of attribute-based access control.

1 Subject-specific attribute

2 Object-specific attribute

3 action: action on resources such as read, write, update and delete.

Capability-based Access Control: Capability-based Access Control[16] is also a DAC model. A row-type view is shown, and each row can be assumed as a capability. DAC and MAC are for individuals, not for groups[17] .

Limitation of all above access control Mechanism: When the number of policies increases, it is difficult to manage all policies. Therefore a flexible access control Mechanism is required which can handle all policies and situations. Role-based access control is static [17] which can't work dynamically. Attribute-based access control can manage such situations, but a large system requires many attributes. when attribute increases number of policies also increases. it is difficult to manage the system. These traditional access control Mechanisms are limited to granted or denied in dynamically changing contexts. In all conditions, they do not provide adequate functionality in distributed and cloud and fog environments [17] .

Zou et al. [18] have improved Context-aware Role based Access Control (CRBAC) model [19] ; especially, for Smart Home setup. In a Smart Home environment, Access to any object of service is time-dependent or activity-dependent. On the usage of the object, this access permission may be denied or granted to limited rights. There are two processes. Attribute updating and Status updating. Attribute updating denotes updating the access control context of each entity and mappings of each entity. For example, the user access control attribute is the user's place, which changes as soon as the user moves. There are so many users which have role's value. Users can modify this value by changing or leaving the role value. It Depends on behavioral constraints which are being satisfied. A session is created with status enabled, that allows users to access permission with the role. But, as soon as any of the constraints is altered in the behavior domain, the session's status check is done and if any constraint is not fulfilled, the session is updated to be disabled. This projected model has the following limitations. It is a centralized model. Therefore a single point of failure can occur, and it increases the communication operating cost. In case of conflicts between attribute values and/or approval, the accountability is set for the administrator. It is supposed that the administrator cautiously plans all the constraints, so it does not have any conflicts in the system. For a usual user who wants to plan his access control rights for this Smart Home, this assumption is tough to accomplish. Context values will have to be predefined by the administrator and updating the values dynamically is unfeasible.

Hernndez-Ramos et al. [20] have made better the capability-based access control model presented by Gusmeroli et al. [9] . They have presented to use distributed architecture instead of given centralized architecture by Gusmeroli et al. [9] . When a user sends a request to access the device. The device itself analyses and grant or denies authorization accordingly. This method gives end-to-end secure access by removing intermediate entities. The key component of a capability-based access control model is the Capability token. This token is a permission to access the entity from the token possessor. The presented technique has four steps. The first step is to give the capability token. This token is given by the device owner to the user who will request to access the device. It is supposed that the token provider and the receiver have already completed mutual authentication by another source. If the user has a token, he sends a request with a token to access the device. This request is sent as a CoAP [21] request, where a token is sent as a component of a payload. On one occasion the device finds the access request, the device performs the authorization process. We found the following limitations of this method. It is supposed that all the devices have resources to execute and store Policy Decision Point (PDP); however, this is not possible in large IoT systems. On one occasion, the token is allotted to the user by the device owner, a user can access the resource several times in a defined time-span. This can consequence in replay attacks while an attacker can just get the token and keeps on replaying it. Generally, a token is validated for a timestamp. In a distributed atmosphere, handling a single clock is a very tough job. This state can consequence in a genuine access request being rejected or even an expired request being permitted. It is supposed that context constraints are known to the device only and therefore they are computed while accessing the request, while there is the possibility that one device may depend on the current context of another device and context information may not be easily found to the device.

Fotiou et al. [22] have presented access control for the IoT resembling the CapBAC model. They proposed a trusted third-party component access control Providers (ACPs). It is between the access requestor and the resource. All the access control decisions are managed by ACPs. They have suggested that the access policies are stored on each resource. When a client sends a request to the device for the resource, the device generates a token and sends it to the client. On receiving, access policy and token, the client forwards them to ACP. ACP analyses the received policy and decides on accepting or rejecting the access. On accepting the access request, the ACP computes the secret session key with help of the received token and client identifier, which is same as what the device calculated. ACP sends this secret session key to the client. The client uses this received key to communicate with the device to access the required resource. We found the following limitations of this method: Given method does not consider any contextual information. It is not feaseble to store complex policies at each resource. The resource also stores many requesters with capability tokens while waiting for the access rights decision has been taken by ACPs. When there are many requests for the same resource, this may result in a communication bottleneck at the resource level due to lacking storage.

Atlam et al. [23] have presented an Adaptive Risk-Based Access Control (RBAC) model using eXtensible Access Control Markup Language (XACML). They presented to build on the XACML access control model to factor in available risk. The risk value is obtained with help of user contexts, risk history, resource sensitivity, and action sensitivity. The access request is either

accepted or rejected, it depends on the evaluated risk value. We found the following limitations of this technique. There is the additional part of the risk factor analysis; the model also suffers from traditional drawbacks of Attribute-Based access control models. It is a centralized model and so a single point of failure may arise along with increasing the communication operating cost. It is possible to manipulate only a single constraint for risk computation and alter the access control decision. When the access request is approved, there is no enforcement on how long the access should be permitted.

Uriarte et al.[24] presented an Expressive Policy-based access control for resource-constrained devices. They presented to usage a model alike to the ABAC model. In this model, the access control policies are saved on the central access control Server. Primarily authorization is performed by access control Server containing calculating policy implementations. If access is accepted, a ticket is issued to the subject. access control Server sends all the policy requirements to the resource for every access request. These consequences high communication operating cost in case of a large number of access requests for a specific resource. When there are two true requests are sent simultaneously for the same resource, it may consequence in one request being rejected on account of local policies are stored that of another request. When context information is sent from the access control Server to the resource and the actual request from the subject, it may generate delay and there is a chance that context value has already been altered.

Schilit et al.[25] have classified a context into three classes based on where you are, whom you are with, and what resources are nearby. The life cycle of a context[26] contains Context Acquisition, Context Modelling, Context Reasoning, and Context Distribution. Context Modelling is the prime requirement of any context management system because it describes how a context will be characterized in the system. Multiple context modeling methods are in the literature that is categorized into Key-value pairs, Mark-up schemes like XML, Graphical models like Unified Role Modelling (UML), Object-oriented models, Logic-based models, and Ontology-based models[27] . Among these context modeling methods, ontology-based modeling is the most suitable method to manage context[28,29]. An ontology is a formal unambiguous explanation of a domain. An ontology is comprised of classes that show entities found in the domain[30] . The basic ontological representation primitives are concepts, relations, instances, and axioms[31] . New axioms are logical results of the represented axioms. these can be derived with help of automatic reasoning using ontologies. Ontologies are subjective to the domain, and therefore, based upon the user's inclinations and context ontology, the quantum of information provided to the user can be reduced. Ontologies are typically defined using Web Ontology Language (OWL). Reasoning in Ontologies can be generated using OWL-Description Logic (OWL-DL) different from OWL.

In 2018, A federated Capability-based Access Control(FedCAC) mechanism has been proposed [6]. Which is a combination of a centralized and decentralized approach. In this model, a local coordinator is available to verify any user to authorize. In this, the context concept is also introduced in providing authorization.

Xu et al.[7] have suggested Federated capability-based access control which has capability management, delegation authority, and access right validation. Authorization is being provided by a third party(coordinator). Which is the token-based approach. FedCAC system consists of PDC policy decisions center at a central server, local coordinator, and service provider. First, all devices are registered and PDC makes coordinator policies and access rights where local entities' databases and related rights are stored. devices which are registered in PDC, whose local devices information is also available at coordinator. Coordinator and PDC are synchronized at a duration. When any device or user wants to request service, its details are reached at the coordinator. The coordinator generates a token; this detail is sent to the service provider. The service provider checks its details and validates them. After that, it grants or denies the service. These are the following limitations in this model. The local coordinator does not have the right to make its policies for authorization. The authorization process has communication between the requester and coordinator after that requester to the service provider. It is increasing communication overhead. The Resource owner's role is only for validation, otherwise, whole authorization is performed by the coordinator i.e., a third party is providing validation. This coordinator may have lots of devices for validation which can generate the same problem of a bottleneck. The concept of context is being used in authorization, but no clear process of authorization computing with the help of context has been proposed in this model.

Mehta et al.[10] have proposed a decentralized context-aware access control model for the IoT. This model contains context reasoning, context updation, and context resolver at every device. It has one central device to get any context value that is not available at the local device. Context resolver resolves any conflict related to context. The system administrator defines an ontology for every device. If any local device does not have a context value, the central context manager can provide these values to that device. If the central context manager does not have a requested context value, it sends a request to another device to find the context value. This device sends context value to the central context manager and the central context manager sends this value to the local device which has requested it. Here device is a service provider.

These are limitations of this model. In this model, local devices do not have the right to define their policies while the local device is more aware of local conditions. For example, if door status information is not available at the local device, so it requests a central device and if the central device does not have this value, it requests it to another device. This device sends the door's present status to the central device and the central device sends it to the requested local device. During this transmission someone changes door status, there is no arrangement of decision revoking.

In the present time, it is required granularity in access control so we used context-aware access control.Context-Aware access control is any information about the access control specific entities, user including resource owner, data resources or environment context-aware access control work in a dynamic environment where users can access anywhere and anytime access of service. Context-aware context information can be classified as a useful resource and their environment-specific condition[32].

Kayes[17] has also provided a similar type of access control Mechanism Based on context-aware access control using fog computing. Fog computing is an intermediate layer between the cloud and IoT devices[33]. Local fog devices are connected to IoT devices and central cloud devices. Fog is a middle layer between clouds. Any clear-cut method or algorithm of authorization issuing has not been provided in Kayes[17].

Above both mechanisms, whether access control decisions are being taken locally but when context information is not available at a local device, they are directly going to the Central device to fetch context information. Therefore it is the dependency on the Central device, so the decentralized concept of authorization is being affected and the last device means the requester must wait for a long time. Its time duration is long, it may be an arrangement of sleep mode but not deep sleep mode because transition time from deep sleep to active mode is long[34]. When any device is idle for a long time it should be chosen deep sleep mode. Therefore deep sleep mode is not applied for seconds or mile-seconds. Sleep mode reduces around 20 percent current requirement[35]. Still waiting time of the local device is so much that it is using more energy consumption than a single device in a decentralized system. Thus, it reduces the lifetime of the battery. IoT devices are installed in remote areas where it is unable to change battery frequently. Our main contribution in this paper is that we are providing mechanisms that not only provide context-aware access control simultaneously; it provides a low active time to Controller. This is the right fact that communication required more battery than computing. But this paper did not calculate communication battery requirements. In the future, we will consider battery requirements in communication.

In this paper, we have proposed a new model of context-aware access control which is increasing the performance of networks and reducing the battery requirement of the Controller which has sensors and is connected to service provider device. Local device(controller) Administrator or Owner is making its policies. Therefore it is not depending on the central device for policy. It can share context information from time to time with the central entity. This thing is out of our scope.

3 | PROPOSED MODEL

We have made this paper which tries to minimize the weak points of the above-described papers, specially Federated capability-based access control[7] and Decentralized Context Aware access control Model for Internet of Things[10]. We have proposed DCAAC model, where Administrators formulate access control policies according to local conditions. Alessandra et.al[36] have said that we should make policies for who can access resources instead of who can't access resources because people who are not authorized to access are more than unauthorized people. It is the concept to reduce the number of policies. To reduce the size of policies that are made to access devices. Policies for access control are defined who is allowed or not because authentication has already been performed. In our experiment we are defining the policy that which group is not allowed.

In our paper, we are designing the ontology and system with the smart home in mind. access control is provided in it according to social relation.

3.1 | System Design and Algorithm

In a centralized system, all decisions are taken on the server without any consultation with the local device. The device must implement those decisions. This model has been shown in Gusmeroli et al.[9]e. In this PDP is policy decision point which is on the server. figure 2 is showing this model. We consider that microcontroller can work as a gateway and be attached to a device.

In a centralized system, multiple devices simultaneously send access requests to the centralized server. Figure 2 shows all the steps that followed for a user request. The communication overhead is high and there are possibilities of security attacks.All these problems are solved automatically in a decentralized system.

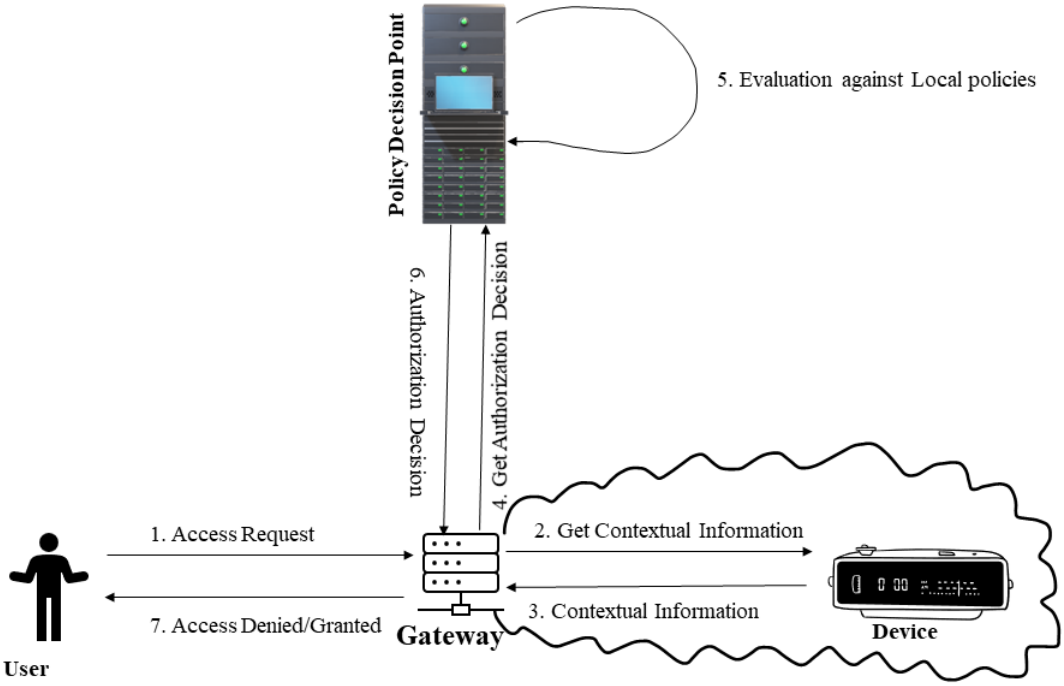


FIGURE 2 Centralized and contextual approach

Figure 3 is showing, Decentralized Context aware access control System Architecture.Its main parts are as follows: input module
device ontology
Temporary ontology + policy generator for context
Reasoner
inference
Policy
Reasoner
In our Experiment, we have used one microcontroller, ontology is stored on this microcontroller. We consider every device has a microcontroller. This microcontroller is connected to the internet and sensors. Sensors provide subjective, objective, and environmental information for decision-making. Policies are coded on the microcontroller in the form of ontology. This is a policy administration point (PAP). Context is handled at a place with help of subject, object, and environmental information it is called Policy Information point (PIP). The interface between the User and Internal components is called the Policy Enforcement point (PEP). Last, the decision taken at that point is called Policy Decision Point (PDP).
We are using two reasoner here. The first reasoner synchronize between the temporal ontology and the device ontology. The second reasoner makes decisions based on the policies from the inference generated by the first reasoner.

In Decentralized access control model based on context Administrators formulate access control policies according to local conditions.The Ontology for context presentation has been defined at microcontrollers and reasoning has been performed on the local device. The microcontroller is a part of the local device. This reasoner gives inference. This inference is used in processing based on policies for decision-making. It is Policy Decision Point. After that permission is granted or denied to access the device
.

python code for Input Module:

```
def entry_page():
    print("This is entry page : Enter the Person characteristics")
    print("Enter Behavior True or False")
    value = input()
    print("Enter Background Good or Crimnal")
```

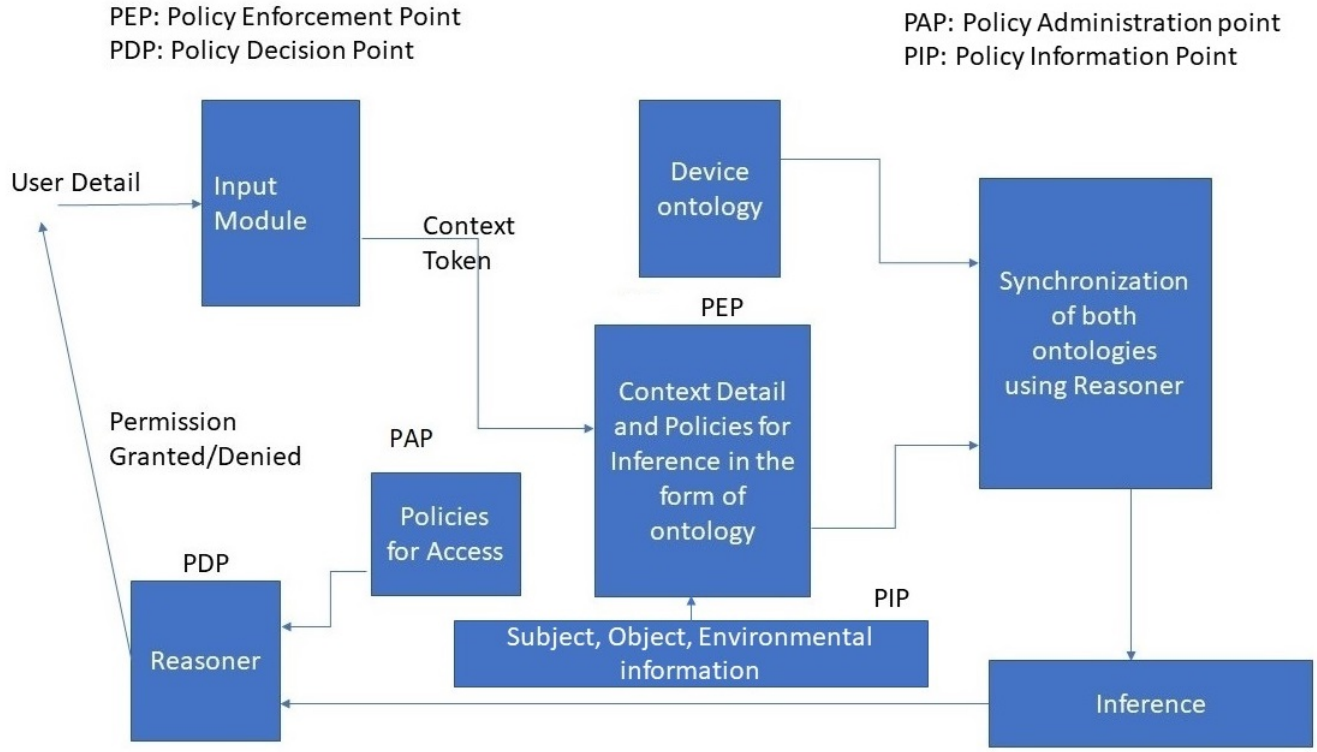


FIGURE 3 Decentralized Context aware access control System Architecture

```
    value = value + " " + input()
    print("Enter Groupings: Single or Married or Villager or NearRelative or FarRelative")
    value = value + " " + input()
    return value
```

python code policy Synchronization and ontology reasoner

```
    onto = get_ontology("Person.owl").load()
ONTO_ID = 0
def page_result(value):
    global ONTO_ID
    ONTO_ID = ONTO_ID + 1
    onto_tmp = get_ontology("http://tmp.org/onto_%s.owl#" %ONTO_ID)
    Behaviour = value[0]
    Past = value[1]
    Groupings = [value[2]]
    with onto_tmp:
        person = onto.Person()
        if Behaviour == "True": person.Behaviour_positive = True
        elif Behaviour == "False": person.Behaviour_positive = False
        if Past:
            Past_class = onto[Past]
            person.has_Past = Past_class()
        for Grouping in Groupings:
            grouping_class = onto[Grouping]
            person.has_grouping.append(grouping_class())
        close_world(person)
```

```
    sync_reasoner([onto, onto_tmp])
class_names = []
print(person.is_a)
for person_class in person.is_a:
    if isinstance(person_class, ThingClass):
        class_names.append(person_class.name)
class_names = "," .join(class_names)
rsl = class_names
onto_tmp.destroy()
return rsl
```

When users are not coming, it can go for deep sleeping mode. And if again any user comes it can be a hardware interrupt arrangement to come in an active state.Figure 4 is showing Decentralized access control model based on context and flowchart of this algorithm is being shown in figure 5.

We are considering that the user is already authenticated by other authenticated means such as Biometric, RFid, etc. device authorization ontology is made according to device nature. For example, if the device is Refrigerator, its domain of access is person and range is Near relatives. If the device is the first door of the house it's the domain of access to all persons. The ontology for every device is different. In a decentralized system, the device is the object which is being used by the User. The request is sent by the user to access any device. Therefore the device owner should have the ability to define access policies according to the local environment, the user, time, and other contexts.

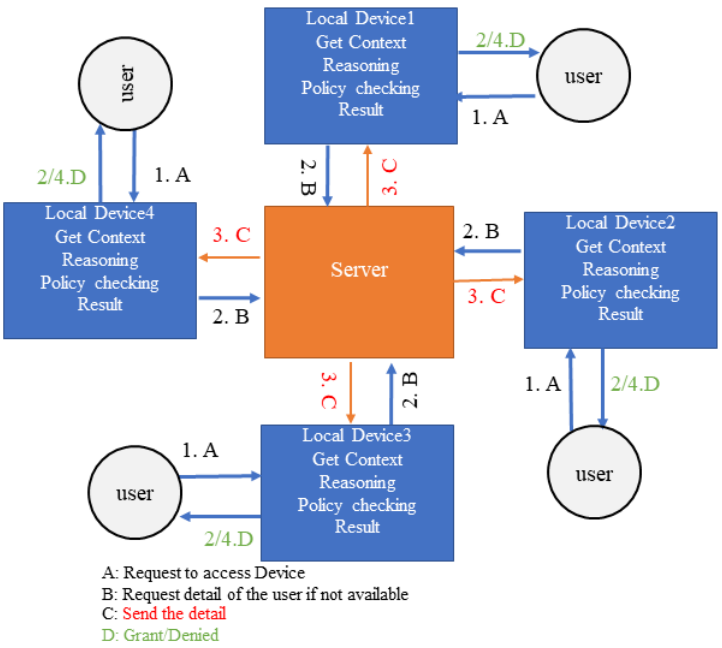


FIGURE 4 Decentralized context-aware access control model

3.2 | Description about Algorithm

This algorithm is for the DCAAC system .

1. Experiment starts with one user request to access services from devices.
2. device gets information about the user's location, time, and family relation context.

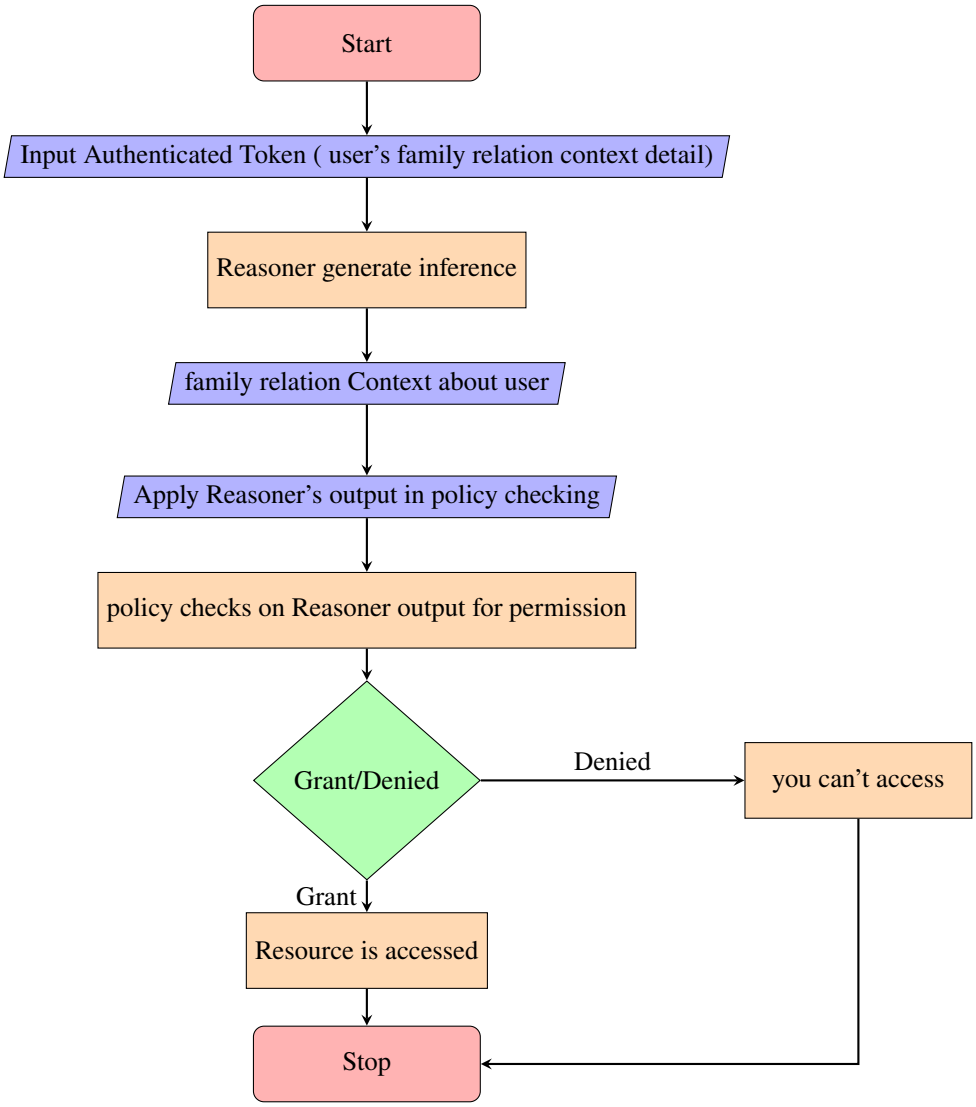


FIGURE 5 Algorithm of context-aware based decentralized access control model

3. device start reasoner to process on the device.

4. Based on the relation between classes and properties in the form of ontology, the reasoner generates inference about the person(user).

5. This inference is an identity token based on context.

6. If this satisfies to policy to access the device, service is provided otherwise denied.

3.3 | Ontology Design

We made ontology using python with the help of the book Ontologies with Python[37]. We made three classes. Grouping, Past, and Person. Figures 6 describes about it. Figure 7, 8, 9 and 10 are describing about sub classes of Person Classes.

When user wants to access any device, request is sent by user. After that reasoner runs and use user's information. This reasoner gives output and reconfigure ontology in the form of inference. This reasoner can generate the following result. Unknown, friend, neighbor, and other. Figure 6,12 and 13 show asserted model and inference model of ontology.

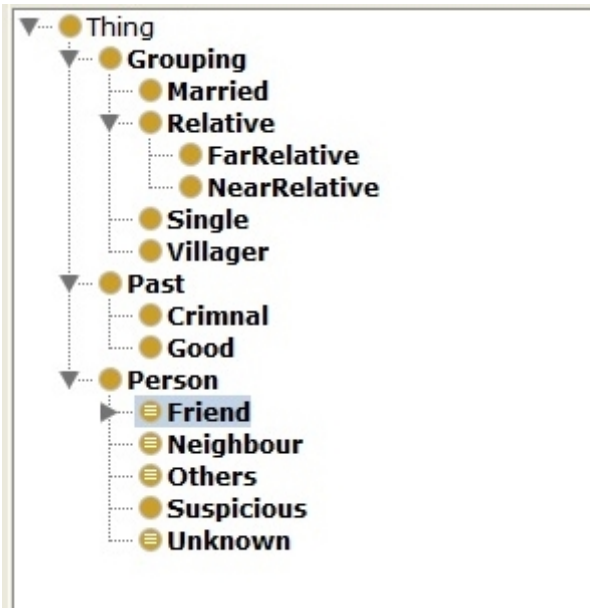


FIGURE 6 Relation based ontology for authorization

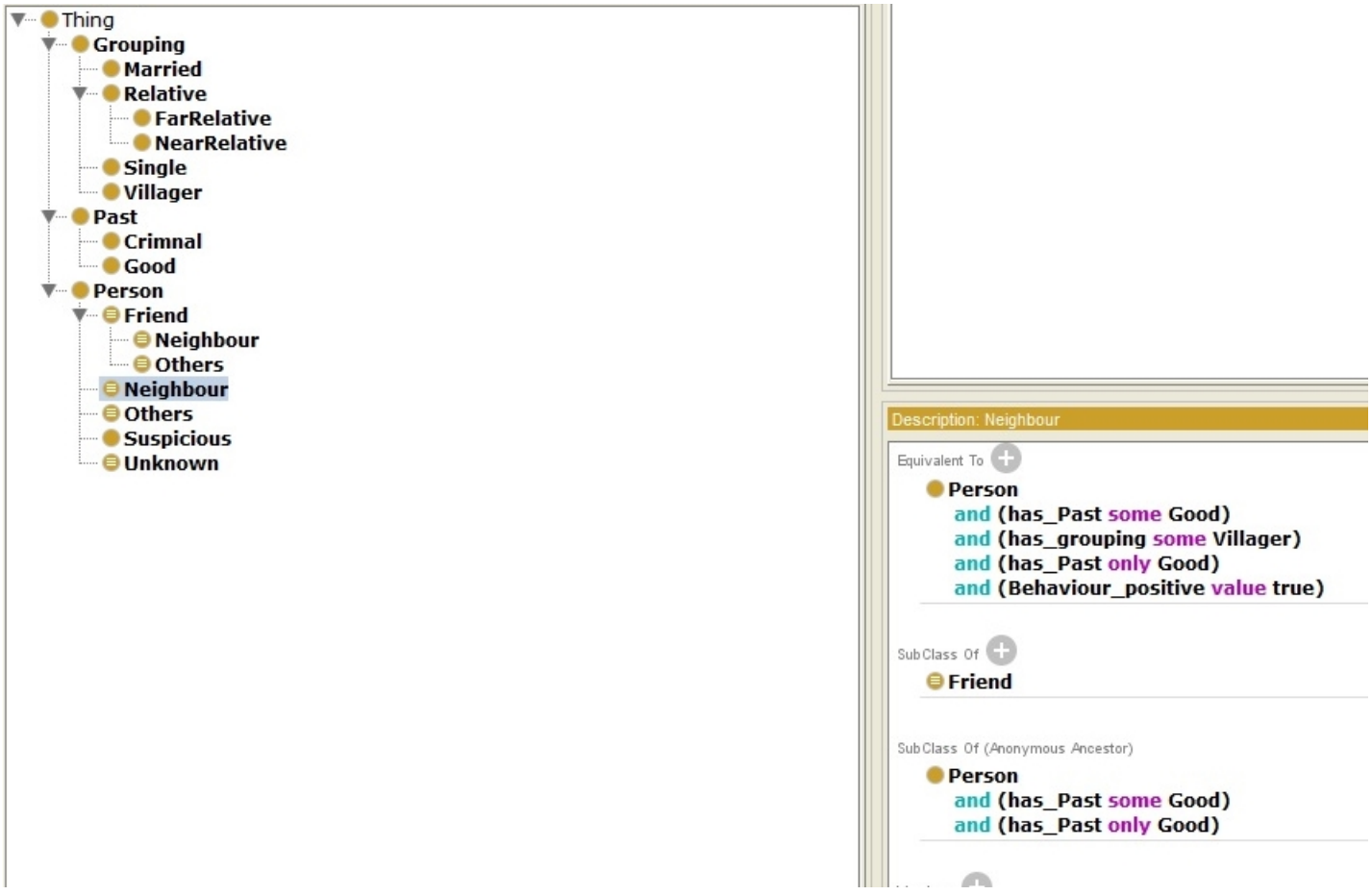


FIGURE 7 Definition of Neighbour class

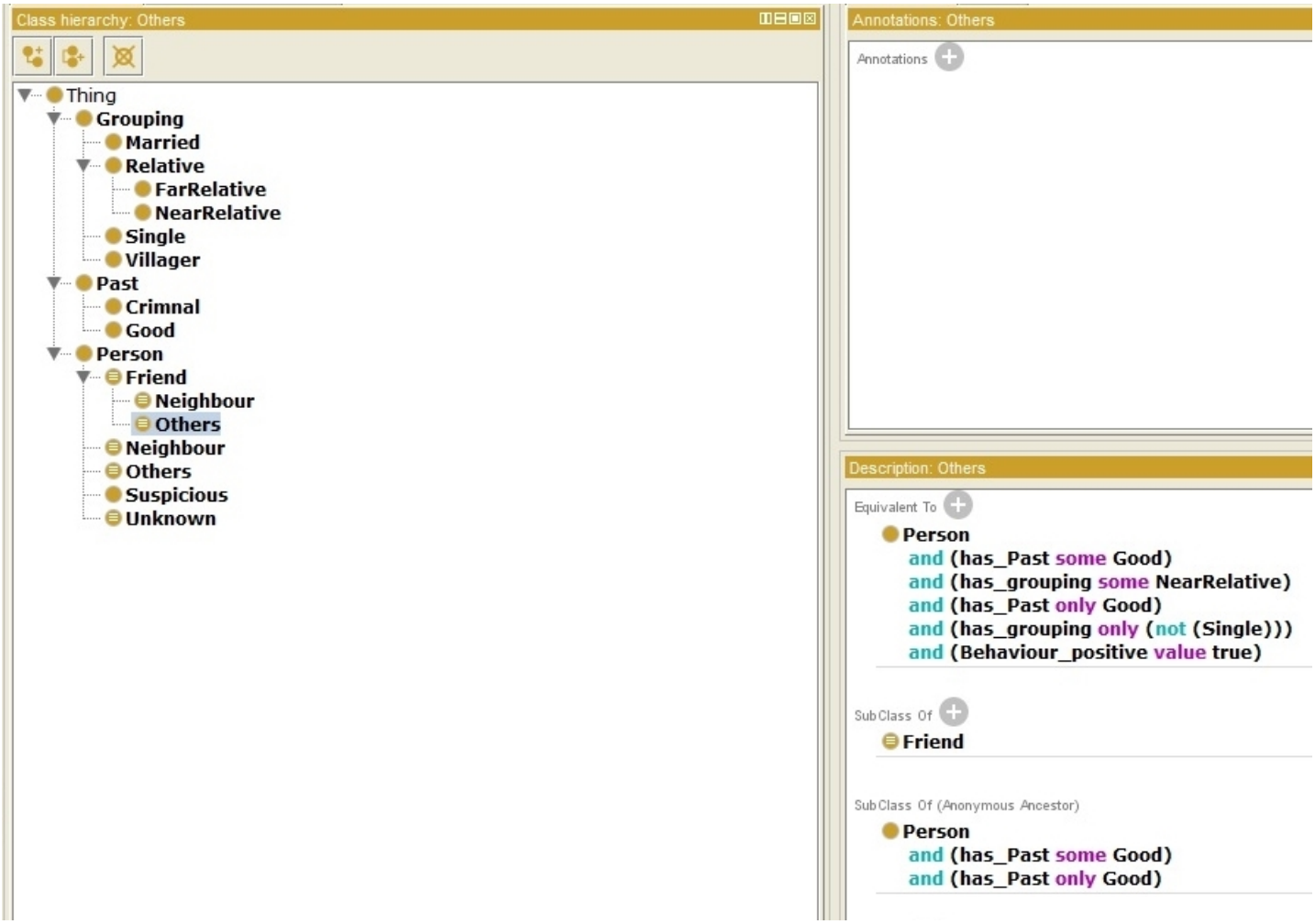


FIGURE 8 Definition of Others Class

We make a policy that an unknown person is not allowed to access any device. If an unknown person tries to access It can generate an alarm. It is getting inference that suspicious is also unknown. Therefore it should not be permitted in the house. Following is its ontology.Figure 8 is showing this relation.

Ontology is used for context presentation. An ontology is a description of things, relationships, and their characteristics, usually in a well-bounded domain[38] . Its reasoner gets inference from the description of things and properties of objects and data, the functionality of relations. In IoT lightweight ontology[39] is used. At present time, in smart homes, granularity in access control is reached to house's room level[11] . In which access should be provided based on public room and private room. This is also operating in a centralized manner. Granular access controls define who can have access to each part of a system, as well as what they can do with that access. However, setting up permissions for each user is impractical and would be incredibly time-consuming to track and maintain[40] .Our access control framework is performing access control in a decentralized manner. We made ontology using python[37] and updated ontology using protege[41] .

We have performed a simulation of a context-aware decentralized access control model and centralized access control, model. In a centralized access control model based on context, users want to access a device. He/ she sends a request to the device, the device forwards the user's information to the server. The server has policies and access rights to detail about users. The server runs reasons and input is detail about the user. Reasoner gives output. this output is used in policies. If policies checks are done for this output. If policies are allowing this user, then a grant signal is sent to the device otherwise denied signal is sent to the device. Based on policies, this device provides service or does not provide service. we obtained that the active time in this model is high if the device does not go to sleep mode.

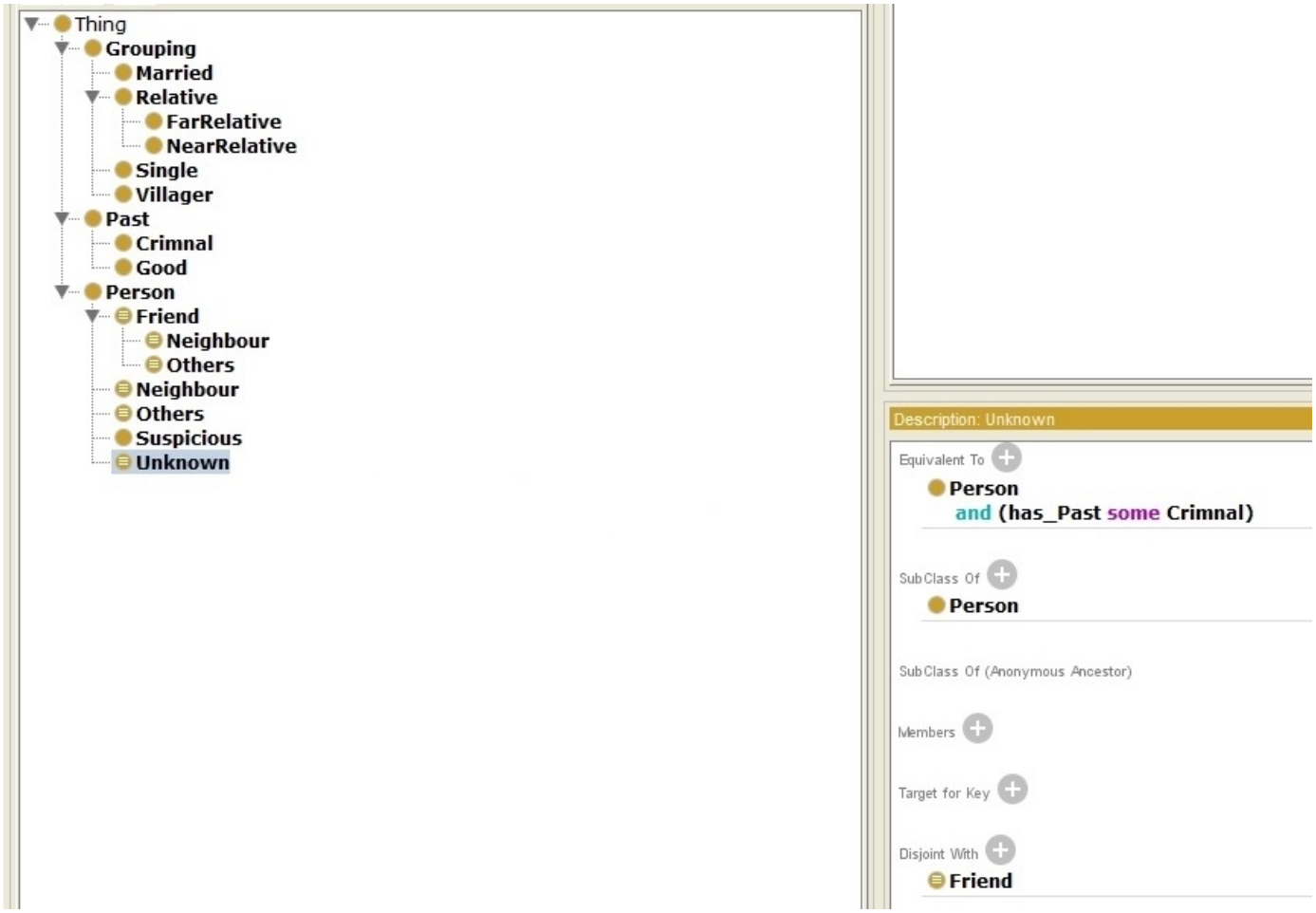


FIGURE 9 Definition of Unknown Class

In decentralized context aware access control , where access is being provided based on family relation context. The user wants to access any device, he/she send a request to the device, then the device starts the reasoner based on user's context information. This reasoner provides output. This output is applied in the policy checker. Policy checkers determine using policies that user is permissible or not. After that service is provided or not. In this, response is fast that is active time is very low and after completion of all jobs local device can go in deep sleep mode.

4 | SECURITY ANALYSIS IN DECENTRALIZED SYSTEM

4.1 | Adversary Model:

Local attackers—irrespective of technical background—can pose a significant threat to the system by cheating physical sensors to sense incorrect references or disrupt the privacy of others. In fact, likely local attackers such as family members, guests, and workers may have sounder inspiration to bypass access control than unaware distant attackers[42]. Our work studies local hazards and emphases on those with legitimate or illegitimate access to home by non-technical users.

4.1.1 | Case 1:

We imagine a scenario in which the only member of the household can take undue advantage of the rules. For example, we made a policy that no child can turn on the TV without the presence of elders. But at night when the elders are sleeping then he can turn on the TV because in that we only included the context of the presence of the elders in the decision making process.

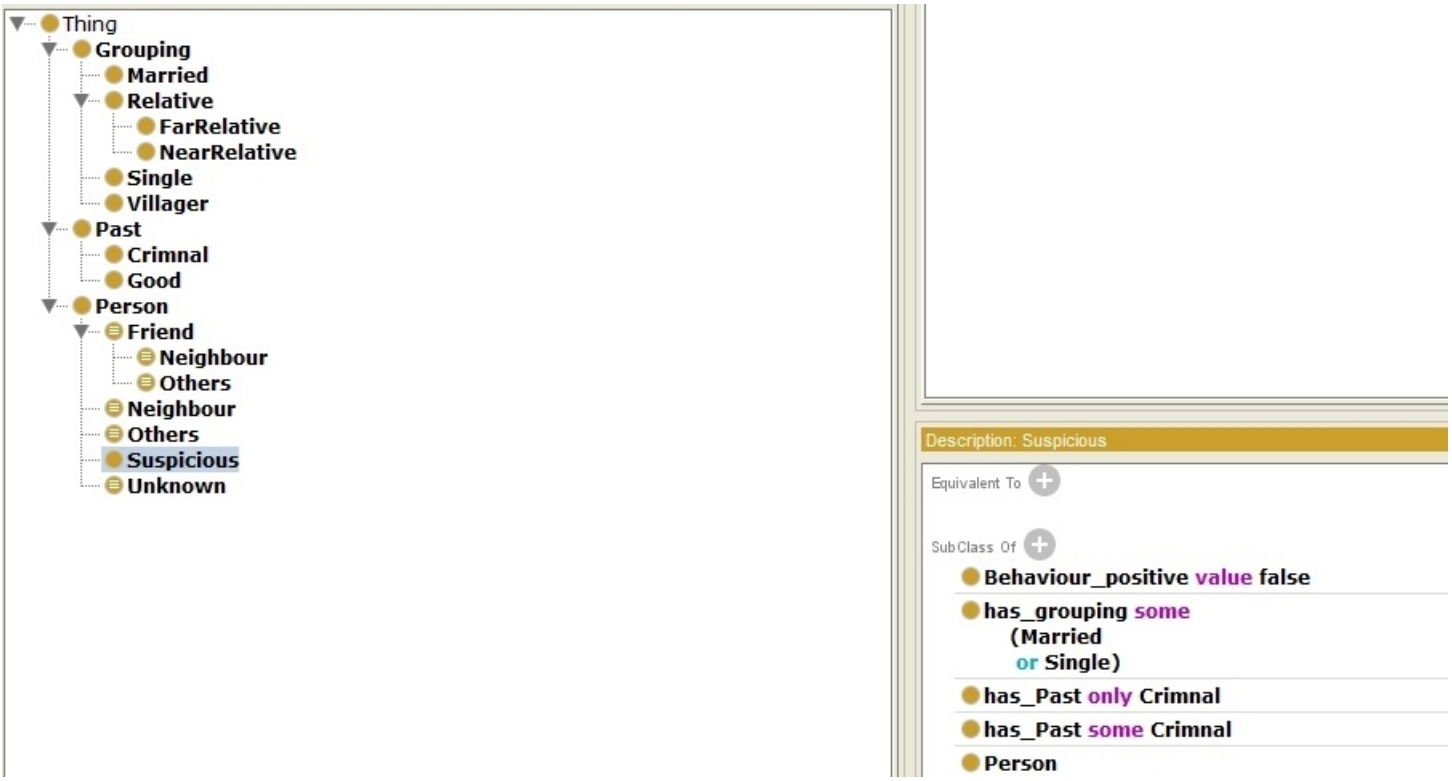


FIGURE 10 Definition of suspicious class

To overcome this limitation, we can also include time in the decision-making process of when the elder is together and at what time he cannot turn on the TV.

4.1.2 | Case 2:

Sometimes based on the results generated through a replay attack, the attacker is able to analyze what situations he can create to gain permission to enter the system. The attacker gains access by providing a suitable environment. figure 13 is showing different types of Threats which are possible in our model.

5 | SECURITY AGAINST ATTACKS

5.1 | Eavesdropping

When communication is taking place between two devices, the third device is listening to it or collects that information, changes it, and sends it again to the destination, it is called an Eavesdropping attack. In a decentralized system, authorization is given by the same device from which the service is to be taken. There is no third entity between the user and the device that can interfere with it.

5.2 | Snooping

It is not possible in a DCAAC. The first phase is authentication. Only authenticated person can enter the system. The second phase is Authorization. Which, we are using the context of time, relation, and location. Which stop snooping.

5.3 | Interception

It is not possible. In a DCAAC, there is no one between user and Service provider device which may intercept information.

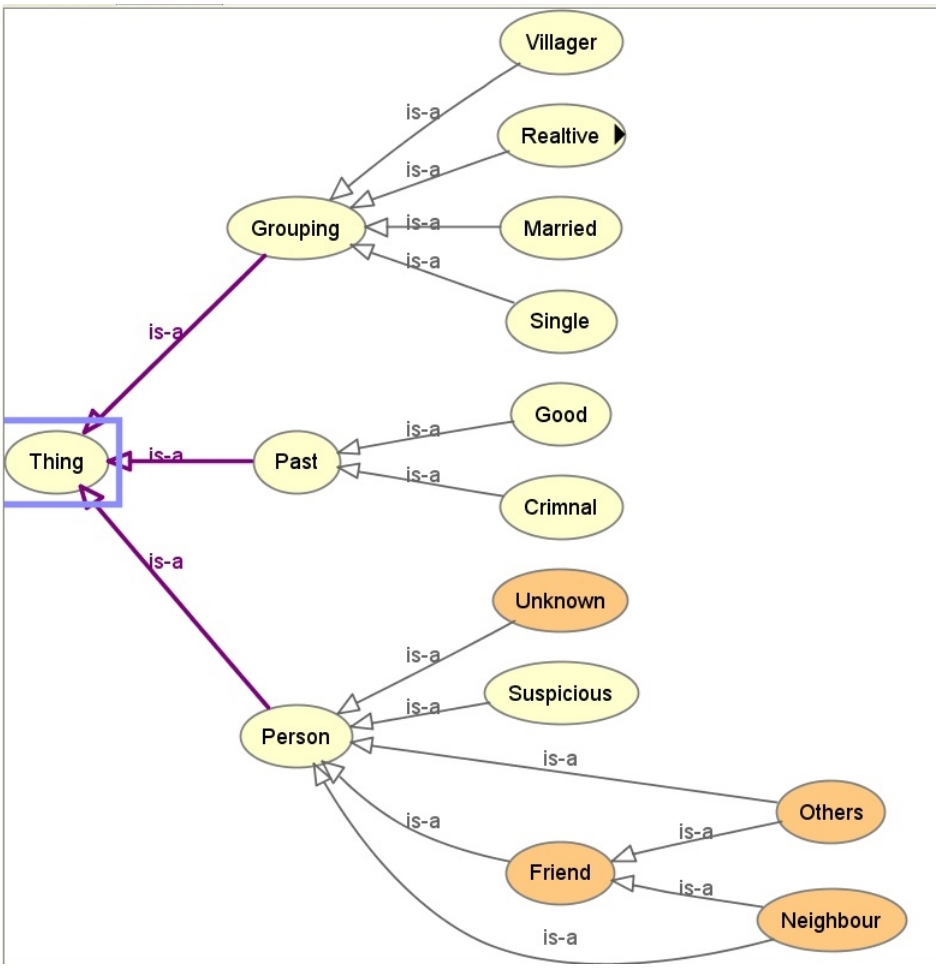


FIGURE 11 Relation based ontology graph1

5.4 | Modification Attacks

It is not possible in DCAAC. User device access is possible under some policies which are formed by the administrator or Resource owner based on contexts such as time, location, and relations.

5.5 | Repudiation Attacks

If anyone performs this attack, the attacker will not find any internal information about the system. We apply the context of location, time, and relation. So only related persons can perform operations.

5.6 | Denial-of-service Attacks and Physical Denial-of-service Attacks

If anyone performs this attack, the attacker must go to the system physically. We have implemented the context of location. First, Authentication is performed after that authorization. Lots of attackers are segregated in the authentication. Remained quantity of persons is not so many they could reach the system in denial-of-service mode.
If someone abrupt communication of sensor and system does not get proper signal, it stops working and this situation is called Physical Denial of Service Attack.

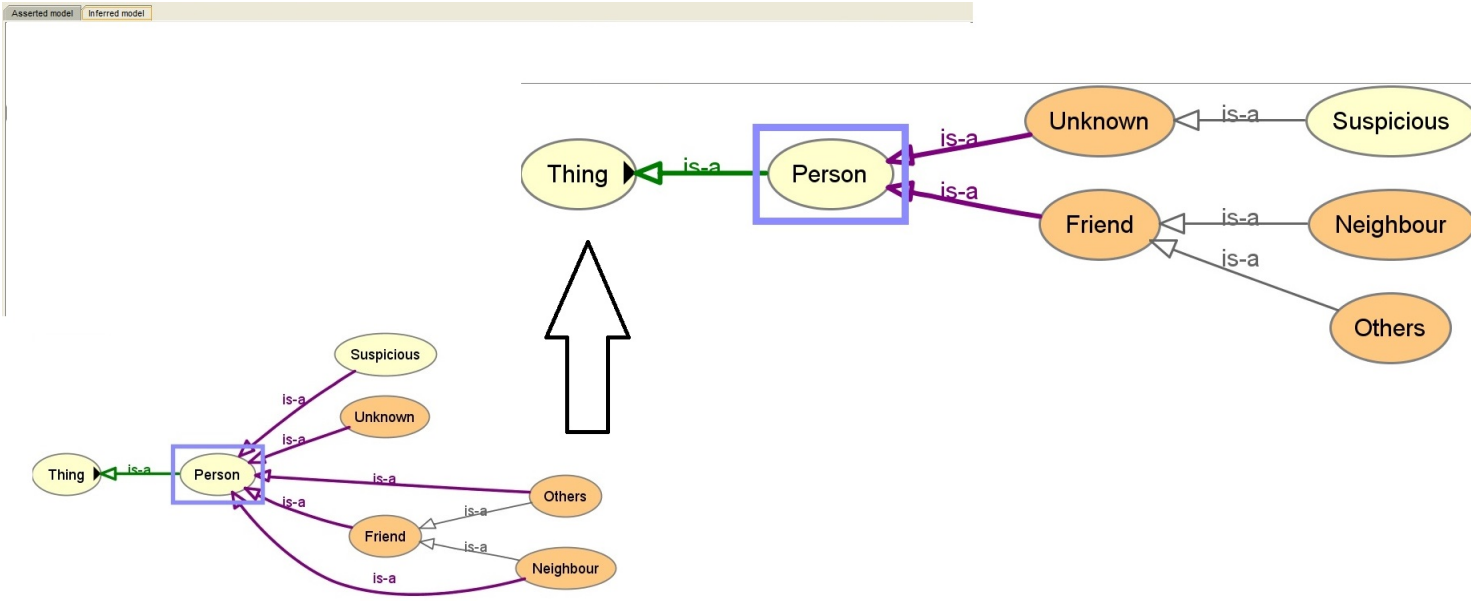


FIGURE 12 Ontology after inference

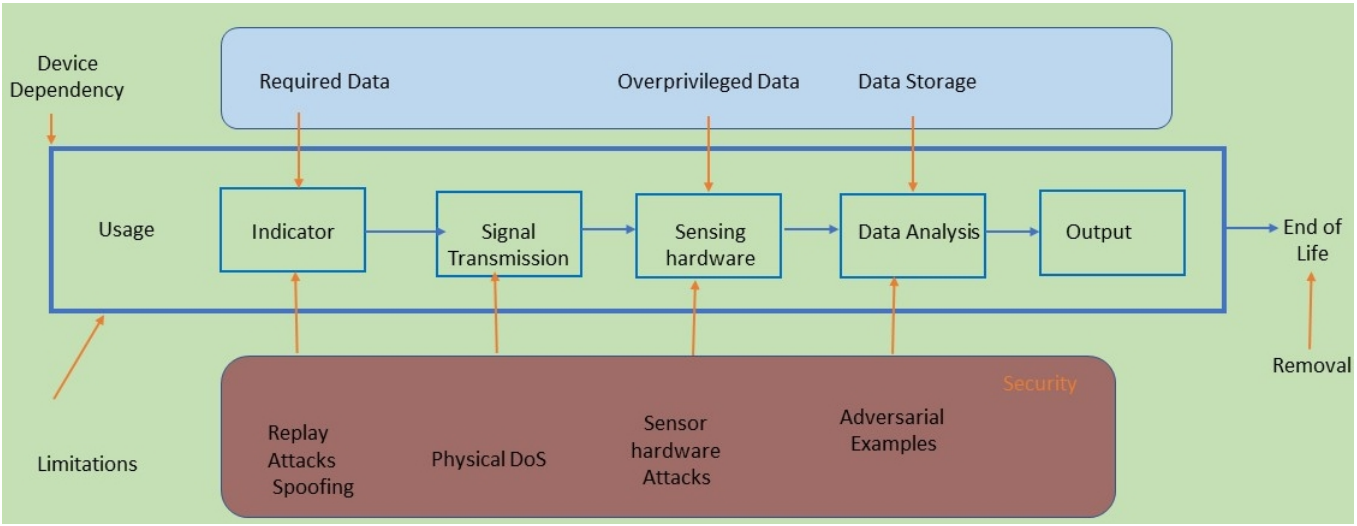


FIGURE 13 Different issues of sensors in Smart home

5.7 | Distributed Denial-of-service Attacks

This is the same reason that to access any system, the attacker must present physically. Where system owner is also there. This attack is performed from multiple sides, but all attackers may go to perform this attack is not feasible.

5.8 | Back door Attacks

This attack is possible when the developer is leaving some flaws in the system, or someone is allowed to enter the system without a login and password. In this system, proper authentication and authorization are being performed using the context of time, location, and relation. Therefore there is no back-door Entry. For an emergency that may be possible.

5.9 | Spoofing Attacks

When anyone tries to make a fake identity. This is segregated by Authentication. For Authorization, only those persons have remained whose identity is true. Therefore our model is considered spoofing proof.

5.10 | Man-in-the-Middle Attacks

In a DCAAC, there is no person between the user and the device which is being used by a user. Therefore there is no possible man-in-the-middle Attack.

6 | EXPERIMENTAL SETUP

Our main aim of this setup to implement DCAAC model based on family relations. Case study is on access control of social and family members. Unknown and suspicious persons should not get access in Home's particular part.
This is machine configuration
CPU:
Architecture: x86_64
CPU op-mode(s): 32-bit, 64-bit
CPU MHz: 3200.000
CPU max MHz: 4600.0000
CPU min MHz: 800.0000
memory 16GiB System memory
memory RAM memory
Topology: 6-Core model: Intel Core i7-8700 bits: 64 type: MT MCP
Drives:
Local Storage: total: 1.03 TiB used: 81.52 GiB (7.8%)
ID-1: /dev/sda vendor: Toshiba model: DT01ACA100 size: 931.51 GiB
temp: 42 C
ID-2: /dev/sdb vendor: SK Hynix model: SC311 SATA 128GB size: 119.24 GiB
temp: 38 C
Python Version:
Python 3.11.0a2+
Hardware :
Arduino Uno.
Budger

7 | ENERGY CONSUMPTION

When current is passed it consumes some energy

$$E = P \times T \tag{1}$$

$$P = V \times I \tag{2}$$

$$E = V \times I \times T \tag{3}$$

where E is Energy Consumption
P is power
T is Time
I is Current

7.1 | Power Consumption Calculation

We are considering that ESP32 with Arduino. followings are technical details for power consumption Transmission time current consumption 240 mile-Ampere. At the time of data receiving current is 95 mile-Ampere. In deep-Sleep mode, current is 10 micro-Ampere. In Modem sleep mode or idle mode current is 4 mile-Ampere[43]. At the time of Computing the Arduino microcontroller consumes 13.92mA[44]. voltage for this experimental setup for wireless sensor network(WSN) is 5V. figure 14 is showing different state of authorization in Centralized access control system..

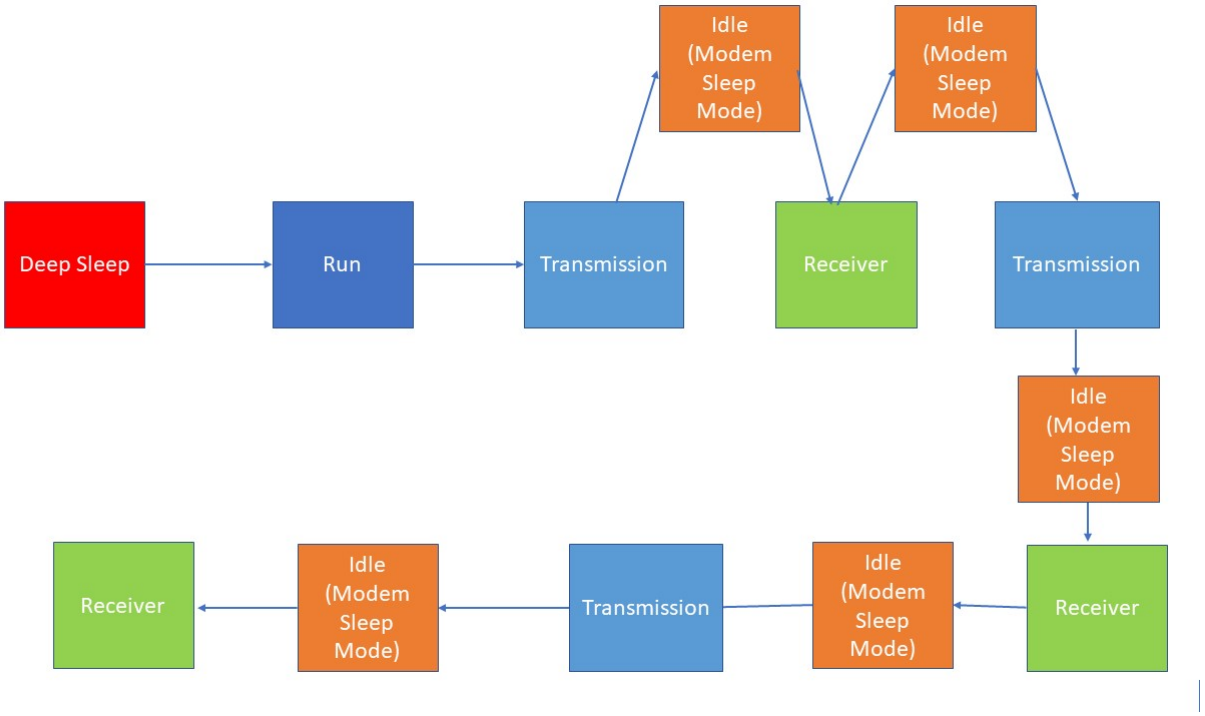


FIGURE 14 States in Authorization Centralized access control System in Centralized System

From Eq. eq 2 Power Consumption in Deep Sleep Mode P_{sl} in Mile-Joule

$$P_{\text{sl}} = 5 \times 0.01 \tag{4}$$

Power Consumption in idle Mode P_{idl} in Mile-Joule

$$P_{\text{idl}} = 5 \times 4 \tag{5}$$

Power Consumption in Transmission Mode P_{tx} in Mile-Joule

$$P_{\text{tx}} = 5 \times 240 \tag{6}$$

Power Consumption in Receiving Mode P_{tx} in Mile-Joule

$$P_{\text{rx}} = 5 \times 95 \tag{7}$$

7.2 | Node and Network Energy Consumption Model (NECM):

In a real-time system, Node consists of a processor, transceiver, and sensor. these all units should be coordinated to help each other. in our experiment, we are considering transceiver and processor. Which participating in performing tasks on node. The processor and transceiver change their state by an external or internal interrupt. For example, external interrupts are commands from other devices. Internal interrupts are periodically data collection.

- Interrupt occurs in the processor component: There are three events when the Processor Energy consumption model(PECM) enters a run state. These are
  1. Sensor generates data, which are collected at regular intervals by the processor
  2. Outsider requests to transfer the data.
  3. When any packet reached the transceiver.
- There are the following states when the Transceiver Energy consumption module (TECM) enters run states.
  1. When PECM starts action to transfer data packet to another node. the 'T_x' enters int run state.
  2. Outside node is sending data, Node has to receive that data, transceiver enter into the 'R_x'

$$E_{\text{sl-run}} = T_{\text{sl-run}} \times (P_{\text{sl}} + P_{\text{run}})/2 \quad , \tag{8}$$

where $E_{\text{sl-run}}$ is Energy Consumption from Sleep to running State. $T_{\text{sl-run}}$ is change from sleep to running state. P_{sl} is power Consumption is sleep State. P_{run} power consumption in running State. Total Energy Consumption

$$E_{\text{total}} = P_{\text{tx}} \times T_{\text{tx}} + P_{\text{rx}} \times T_{\text{rx}} + P_{\text{cmp}} \times T_{\text{cmp}} + P_{\text{idl}} \times T_{\text{idl}} + E_{\text{sl-run}} \quad , \tag{9}$$

$$T_{\text{rx}} = T_{\text{tx}} = L/R \quad , \tag{10}$$

where E_{total} Total Energy Consumption. P_{tx} is Power Consumption in Transmission. T_{tx} is transmission time.P_{rx} power Consumption in Receiving. T_{rx} is Duration in receiving. P_{idl} is power consumption in idle time. T_{idl} is duration of Idle time. $E_{\text{sl-run}}$ is Energy Consumption from Sleep to running State. L is the length of sent or received of the data packet.

Energy consumption in a centralized access control system. Suppose transmission range of side of the square and it is not reaching to the diagonal node. Among them, a single node propagates to the other three nodes till it reaches the base station. Shortest path finding. Once, the path is established, it has been fixed that communication shall be performed through certain nodes. In that situation, single broadcast propagation makes other nodes active who are in the transmission range.

In CCAAC, suppose that there are n links in the shortest path to reach the base station from the requested node. There are 3 trips in the authorization process[9]. In one round trip, one node will activate 2 times. One time of sending and receiving energy consumption is E_{tx} and E_{rx} respectively. Three round trip energy consumption

$$E_{\text{trans}} = 3 \times 2(E_{\text{tx}} + E_{\text{rx}}). \tag{11}$$

We are considering all nodes are equally distributed and when data is transmitted from one node to the base station. It crosses more than one node to reach the base station. We are considering the DSR(Dynamic source routing) algorithm for the transmission of data[45]. The data transmission rate is 1 MBps. Suppose that there are n nodes through which the request is reaching the base station, total energy consumption in transmission, and receiving the request.

$$E_{\text{trans_total}} = n \times 3 \times 2(E_{\text{tx}} + E_{\text{rx}}). \tag{12}$$

Maximum data transmission speed

$$R = 1MBps = 1024 \times 1024 bits/second. \tag{13}$$

Thus, total energy consumption in transmission

$$E_{\text{tx}} = P_{\text{tx}} \times L_{\text{tx}}/R \tag{14}$$

Here, we are supposing that one-time broadcasting is reaching three nodes. Out of them, one node will propagate to the next other three nodes, the remaining two nodes do not transmit because the other two nodes are not in the shortest path.Energy consumption in receiving data is one and half times greater than energy consumption in transmission mode. Because it gets a lot of data, half of the data is out of use[45].

$$E_{\text{rx}} = 3 \times (3/2) \times P_{\text{rx}} \times L_{\text{rx}}/R \tag{15}$$

In our algorithm following Equation is using for energy consumption.

$$\begin{aligned} E =& 2 \times n \times P_{\text{tx}} \times L_{\text{tx}}/R + 2 \times 3 \times n \times (3/2) \times P_{\text{rx}} \times L_{\text{rx}}/R + P_{\text{cpu-state}} \times \\ & T_{\text{cpu-state}} + 2 \times n \times p_{\text{idl}} \times T_{\text{idl}} + n \times E_{\text{sl_run}} \end{aligned} \tag{16}$$

Our simulation is for a one-time authorization process. In Centralized Authorization when anyone wants to access any device, it starts the device from sleep to active state. It sends a request to the resource owner, and it receives an Authorization grant. This Authorization token is sent to the token server and the token is received by the client. again, this is sent to the service server after that service is provided or denied.

In DCAAC, At starting, the device will be in deep sleep mode, when any user wants to access that device, the user sends a request to access it, and it goes into active mode. After that device is in Running state throughout the one time request. Energy Consumption Decentralized System.

$$E = E_{\text{sl-run}} + E_{\text{tx}} + E_{\text{rx}} + E_{\text{cmp}} \tag{17}$$

8 | RESULT

8.1 | Case 1.

When a client is sending a request to the server for authorization. When the number of links are increasing, energy consumption is also increasing.Table 1 and figure 15 are showing result of energy consumption in centralized system.

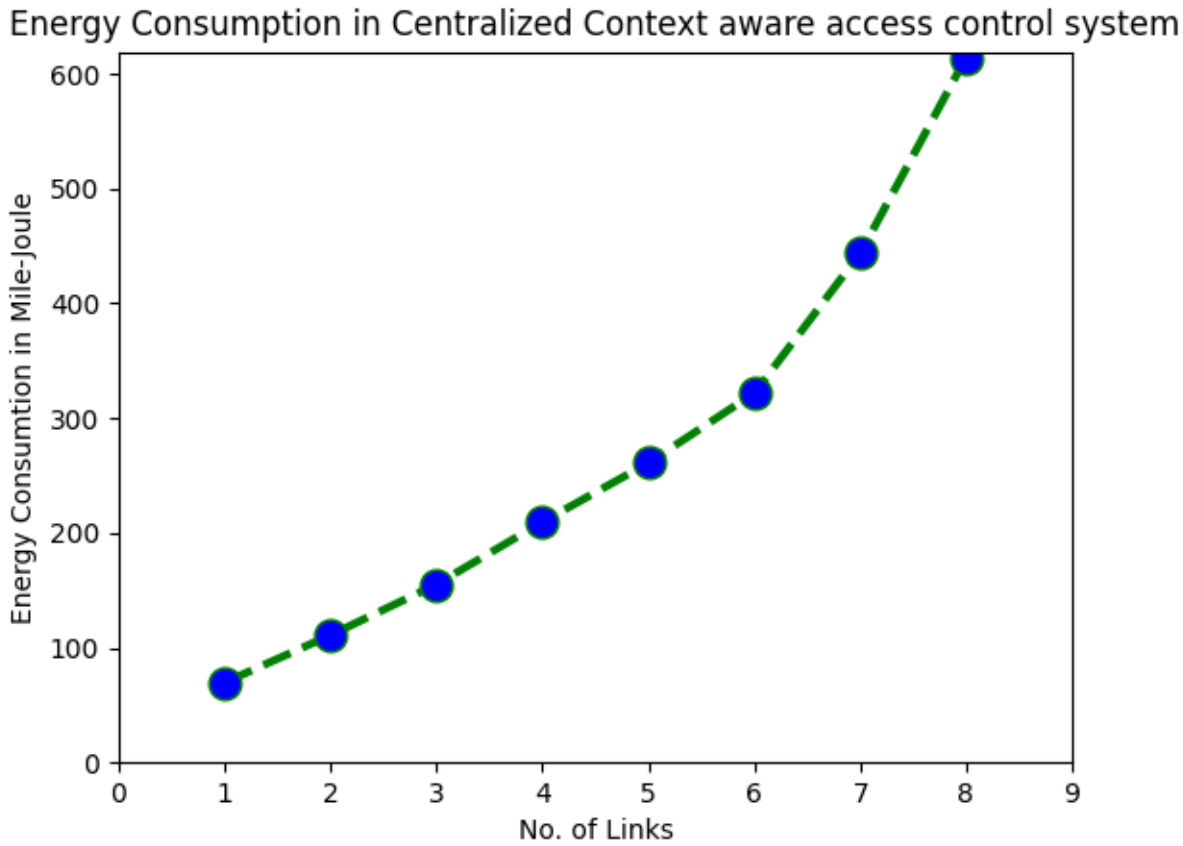


FIGURE 15 Energy consumption in Centralized context-aware access control model

TABLE 1 Energy Consumption(Mile-Joule) per Request in Centralized in Network.

No. of Links	Energy Consumption(Mile-Joule) in Network
1	69.68157002496719
2	110.92624311447143
3	155.1721426963806
4	209.44289426803587
5	260.89855475425725
6	321.67432765960694
7	443.4333686828613
8	614.0964836597443

TABLE 2 Average Energy Consumption(Mile-Joule) per Request in CCAAC in Network.

No. of Request	Average Energy Consumption(Mile-Joule) per Request
1	82.57477164993286
2	96.24131642112731
3	119.97590022811889
4	155.14430619010926
5	163.60036006698607
6	181.48838058242796
7	221.6697367263794
8	270.67603890428546
9	273.9242532325745
10	453.6298492861747

8.2 | Case 2.

When we are sending more than one request simultaneously. It is increasing energy consumption as it is increasing the number of requests for authorization. In this case, the client is directly connected to the server. This is energy consumption for the whole network. Table 2 and figure 15 are showing the result. **In Centralize system** and Number of links between client node and Base Station is 1.
Transmission + Receiving Time = 0.00057220458984375 s
Deep Sleep Time = 0.2948753833770752 s
Computing Timing = 0.006981372833251953 s
Idle time = 0.8780152797698975s
Energy Consumption = 75.50626713037491 mile-joule

8.3 | Case 3.

In Context-aware based Decentralized access control system, we send the request for authorization. For 10 requests for authorization sequentially, all-time about the same energy consumption is found. In centralized context-aware access control, average energy consumption per request is increasing. Table 3 and figure 16 are showing the results.
In Decentralized System:
Total Energy Consumption (in Mile-joule) : 58.294302177429195

TABLE 3 Average Energy Consumption(Mile-Joule) per Request in DCAAC.

n^{th} Request	Energy Consumption(Mile-Joule)
1	58.28965187072754
2	57.037125205993654
3	56.206032371521
4	56.86710548400879
5	55.43575859069824
6	57.87623462677002
7	55.59017715454102
8	54.5914924621582
9	54.359567642211914
10	55.42291488647461

Total Time (Active + Waiting) Time (in seconds) : 0.8340485095977783

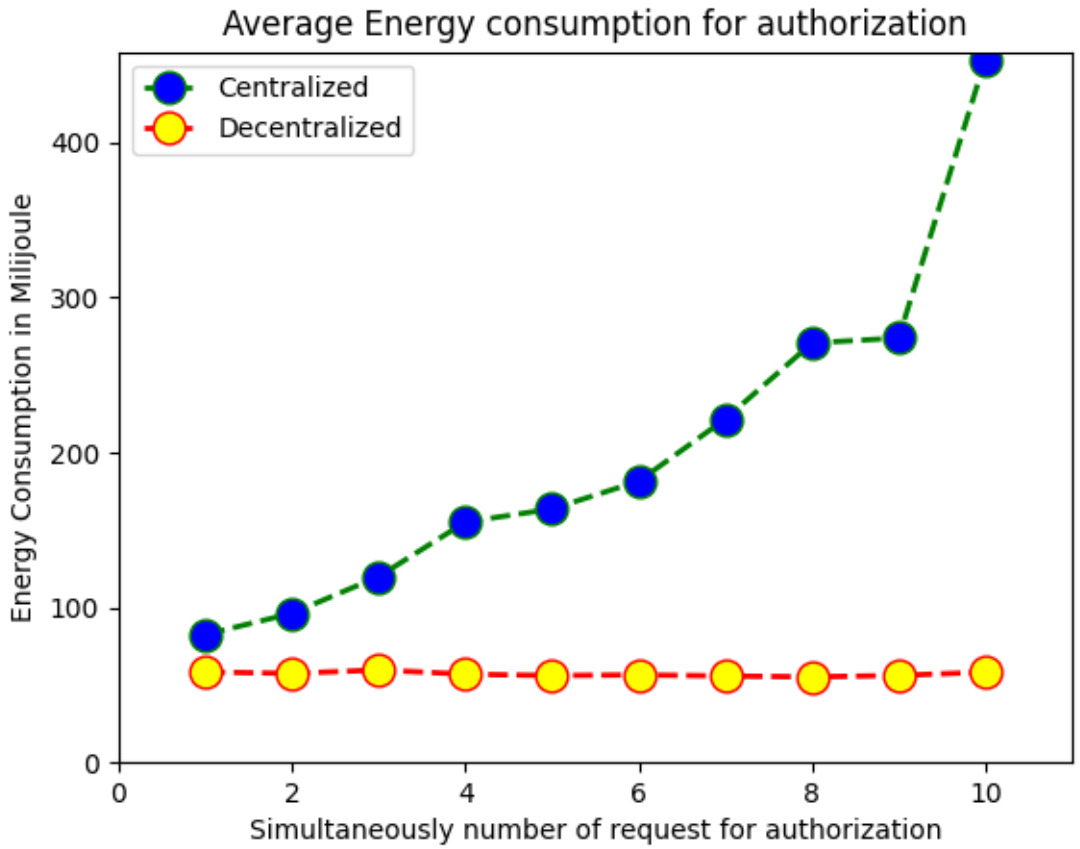


FIGURE 16 Improvement in Energy consumption in context-aware based decentralized access control model

8.4 | Energy Consumption Comparison:

Percentage Improvement in Energy Consumption in decentralized mode than centralized mode is being calculate with help of following Equation. Figure 11 is showing this improvement.Percentage Improvement in Energy Consumption in decentralized mode than centralized mode is being calculate with help of following Equation. Figure 11 is showing this improvement. It is 41.67 % improvement in Energy Consumption of IoT Network in the Decentralized access control system for single request for Authorization by following equation.

$$\%E_{\text{imp}} = 100 \times (E - E_{\text{dec}})/E \tag{18}$$

9 | DISCUSSIONS

At starting, the device will be in **deep sleep mode**[34], when any user wants to access that device, the user sends a request to access it, and it goes into active mode. When the device is working but waiting for a response from another side, it goes into modem-sleep(idle) mode. So that it can quickly come into an active state. Our experiment is showing that an energy consumption centralized system is more than a decentralized system. When the number of nodes between clients and the base station is increasing energy consumption of the network and response and waiting time are also increasing. In a decentralized system, the user is authorized by the same device which is providing services. There is no third entity between the user and the device. In a Decentralized system, Energy Consumption is less, and waiting and response time is also less.

Security is not only important in IoT, but it needs to keep continued devices working for a long time. that is device battery should be for a long time. for a long battery life device, it is required to reduce the active time of the device. after that device could go in deep sleep mode. it can switch to active mode after the interrupt, and it can save energy.

10 | CONCLUSIONS

In this paper, we present a decentralized context-based access control mechanism for IoTs. There are different types of contexts such as context related to the situation of the environment, the behavior of the entity, and relations between entities. In this paper, we considered the context of relations (family relation Context) of entities. We proposed a decentralized context-aware access control(DCAAC) that not only secure the network, but it also increases battery performance, scalability of IoT network. It provides faster responses while improving the battery life of the controller device in IoT. The proposed solution has been implemented as a decentralized access control based on the context in IoT. Administrators formulate access control policies according to local conditions- thereby it allows better and customized policies depending on the situations. Such approach reduces the response time, without sacrificing any bit of the security. Our experiments show that the proposed mechanism saves significant amount of battery power. Man-in-the-middle, Eavesdropping, snooping, interception, and modification attacks, Repudiation Attacks, Denial-of-service Attacks, Distributed Denial-of-service Attacks, and Back door Attacks are not possible in a decentralized context-aware access control(DCAAC).

In future, we will work on the process of discontinuing or continuing the service depending on the nature of the service provided by the device, if circumstances change while using the device.

References

1. Zhang WE, Sheng QZ, Mahmood A, et al. The 10 Research Topics in the Internet of Things. 2020: 34-43. doi: 10.1109/CIC50333.2020.00015

2. Gillis AS. What is internet of things (IoT)?. 2020. https://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT, Accessed 23 February 2022.

3. Surendra T, Anandhi R. Remote Medical Alert Monitoring System - Providing Telehealth By Integrating Ayurveda With Sensors And Mobile Computing. *International Journal of Medical and Health Sciences* 2017; 6.

4. Schuster R, Shmatikov V, Tromer E. Situational Access Control in the Internet of Things. In: CCS '18. Association for Computing Machinery; 2018; New York, NY, USA: 1056–1073

5. Alaba FA, Othman M, Hashem IAT, Alotaibi F. Internet of Things security. *Journal of Network and Computer Applications* 2017; 88: 10-28.

6. Hu D, Kuhn R, Schnitzer A, et al. Guide to Attribute based Access control definitions and considerations. https://nvlpubs.nist.gov/nistpubs/specialpublications/NIST.sp.800-162.pdf, Accessed February 23, 2022.

7. Xu R, Chen Y, Blasch E, Chen G. A Federated Capability-based Access Control Mechanism for Internet of Things (IoTs). In: ; 2018

8. Arakawa J, Sasada K. A decentralized access control mechanism using authorization certificate for distributed file systems. In: ; 2011: 148-153.

9. Gusmeroli S, Piccione S, Rotondi D. A capability-based security approach to manage access control in the Internet of Things. *Mathematical and Computer Modelling* 2013; 58: 1189-1205. doi: 10.1016/j.mcm.2013.02.006

10. Mehta S, Jinwala DC, Dalal UD, Patel DR. Decentralized Context Aware Access Control Model for Internet of Things. *Journal of Computational and Theoretical Nanoscience* 2020; 17(1): 162-171.

11. Dutta S, Chukkapalli SSL, Sulgekar M, Krithivasan S, Das PK, Joshi A. Context Sensitive Access Control in Smart Home Environments. In: ; 2020: 35-41

12. Chen H, Finin T, Joshi A. An Ontology for Context-Aware Pervasive Computing Environments. *The Knowledge Engineering Review* 2003; 18. doi: 10.1017/S0269888904000025

13. Lampson B. Protection. In: ACM SIGOPS Oper. Syst. Rev.; 1974.

14. Qiu L, Zhang Y, Wang F, Kyung M, Mahajan HR. Trusted Computer System Evaluation Criteria. In: ; 1985.

15. Sandhu R, Coyne E, Feinstein H, Youman C. Role-based access control models. *Computer* 1996; 29(2): 38-47. doi: 10.1109/2.485845

16. Sandhu R, Samarati P. Access control: principle and practice. *IEEE Communications Magazine* 1994; 32(9): 40-48. doi: 10.1109/35.312842

17. Kayes ASM, Kalaria R, Sarker IH, et al. A Survey of Context-Aware Access Control Mechanisms for Cloud and Fog Networks: Taxonomy and Open Research Issues. *Sensors* 2020; 20(9). doi: 10.3390/s20092464

18. Zou D, Park JH, Kim TH, Chen X. SH-CRBAC: Integrating Attribute and Status Constraints into the RBAC Model in Smart Home Systems. *The Computer Journal* 2008; 52(8): 861-870. doi: 10.1093/comjnl/bxn046

19. Zou D, He L, Jin H, Chen X. CRBAC: Imposing multi-grained constraints on the RBAC model in the multi-application environment. *Journal of Network and Computer Applications* 2009; 32(2): 402-411. doi: https://doi.org/10.1016/j.jnca.2008.02.015

20. Hernández-Ramos J, Jara AJ, Marin L, Skarmeta A. Distributed capability-based access control for the internet of things. *Journal of Internet Services and Information Security (JISIS)* 2013; 3: 1-16.

21. Shelby Z, Hartke K, Bormann C. The Constrained Application Protocol (CoAP). RFC 7252; 2014

22. Fotiou N, Kotsonis T, Marias G, Polyzos G. Access Control for the Internet of Things. In: ; 2016: 29-38

23. Atlam H, Alassafi M, Alenezi A, Walters R, Wills G. XACML for Building Access Control Policies in Internet of Things. In: ; 2018

24. Uriarte M, Astorga J, Jacob E, Huarte M, Carnerero M. Expressive Policy-Based Access Control for Resource-Constrained Devices. *IEEE Access* 2017; PP: 1-1. doi: 10.1109/ACCESS.2017.2730958

25. Schilit B, Adams N, Want R. Context-Aware Computing Applications. In: ; 1994: 85-90

26. Retima F, Benharzallah S, Kahloul L, Kazar O. A comparative analysis of context-management approaches for the Internet of Things. *International Arab Journal of Information Technology* 2017; 14: 578-585.

27. Yurur O, Liu CH, Sheng Z, Leung VCM, Moreno W, Leung KK. Context-Awareness for Mobile Sensing: A Survey and Future Directions. *IEEE Communications Surveys Tutorials* 2016; 18(1): 68-93. doi: 10.1109/COMST.2014.2381246

28. Perera C, Zaslavsky A, Christen P, Georgakopoulos D. Context Aware Computing for The Internet of Things: A Survey. *IEEE Communications Surveys Tutorials* 2014; 16(1): 414-454. doi: 10.1109/SURV.2013.042313.00197

29. Cabrera O, Franch X, Marco J. Ontology-based context modeling in service-oriented computing: A systematic mapping. *Data & Knowledge Engineering* 2017; 110: 24-53. doi: https://doi.org/10.1016/j.datak.2017.03.008

30. Noy N, Mcguinness D. Ontology Development 101: A Guide to Creating Your First Ontology. *Knowledge Systems Laboratory* 2001; 32.

31. Gómez-Romero J, Bobillo F, Delgado M. Context Representation and Reasoning with Formal Ontologies. In: AAAIWS'11-04. AAAI Press; 2011: 26–31.

32. Kayes A, Rahayu W, Dillon T, Chang E, Han J. Context-aware access control with imprecise context characterization for cloud-based data resources. *Future Generation Computer Systems* 2019; 93: 237-255. doi: https://doi.org/10.1016/j.future.2018.10.036

33. Aleisa MA, Abuhussein A, Sheldon FT. Access Control in Fog Computing: Challenges and Research Agenda. *IEEE Access* 2020; 8: 83986-83999. doi: 10.1109/ACCESS.2020.2992460

34. Gupta S. ESP32 Active Mode and Deep Sleep Mode Power Consumption Comparison. 2021.

35. Arduino Power Consumption Normal & Sleep. 2013.

36. Toninelli A, Montanari R, Kagal L, Lassila O. A Semantic Context-Aware Access Control Framework for Secure Collaborations in Pervasive Computing Environments. In: Cruz I, Decker S, Allemang D, et al., eds. *The Semantic Web - ISWC 2006*Springer Berlin Heidelberg; 2006; Berlin, Heidelberg: 473–486.

37. Lamy JB. *Ontologies with Python* . 2021

38. Powell J, Hopkins M. 5 - Ontologies. In: Chandos Information Professional Series. Chandos Publishing. 2015 (pp. 31-43)

39. Davies J. *Lightweight Ontologies*: 197–229; Dordrecht: Springer Netherlands . 2010

40. The Six Ws of Granular Access Control. https://www.coresecurity.com/blog/six-ws-granular-access-control,Accessed 25 February 2022.

41. Musen MA. The protégé project: a look back and a look forward. *AI Matters* 2015; 1(4): 4-12. doi: 10.1145/2757001.2757003

42. He W, Zhao V, Morkved O, et al. SoK: Context Sensing for Access Control in the Adversarial Home IoT. In: ; 2021: 37-53

43. https://randomnerdtutorials.com/esp32-deep-sleep-arduino-ide-wake-up-sources/,Accessed September 2, 2022.

44. GIANT AT. Reducing Arduino Power Consumption. https://learn.sparkfun.com/tutorials/reducing-arduino-power-consumption/all, Accessed February 23, 2022.

45. Zhou HY. Modeling of Node Energy Consumption for Wireless Sensor Networks. *Wireless Sensor Network* 2011; 03: 18-23. doi: 10.4236/wsn.2011.31003

AUTHOR BIOGRAPHY



Surendra Tyagi.
Surendra Tyagi is a research scholar in the Department of Computer Science & Engineering, Indian Institute of Technology, Jammu.
Research area is privacy and security in IoT and Computer Networks.
M.Tech., (CSE),2014 Delhi Technological University, Delhi
B.E., 2007,MITS Gwalior



Subhasis Bhattacharjee.
Subhasis Bhattacharjee is an associate Professor in the Department of Computer Science & Engineering, Indian Institute of Technology Jammu.
Ph.D., (CS), 2011 Indian Statistical Institute, Kolkata
M.Tech. (CS), 1998 Indian Statistical Institute, Kolkata
B. E., 1996, S.V. Regional College of Engineering, Surat, (now SVNIT).



Devesh C Jinwala
Devesh C Jinwala, Professor in Computer Engineering at the Department of Computer Engineering, Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat, India.
Dean(Research Consultancy) at SVNIT, Surat.
Adjunct Professor in Computer Science and Engineering, at IIT Jammu.

How to cite this article: Suredra Tyagi, (2022), Low-Energy Decentralized Context-aware Access Control in Internet of Things, *Security and Privacy*, *202;00:1–6*.

MAPAN-Journal of Metrology Society of India
https://doi.org/10.1007/s12647-024-00736-3

ORIGINAL ARTICLE

Measurement of Mechanical Properties of Friction Stir Welding of Al–Li alloy under different environmental conditions

S. Singh, N. Yuvaraj* and R. Wattal

Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India

Received: 02 October 2023 / Accepted: 11 January 2024

© Metrology Society of India 2024

Abstract: The butt joints of AA2099-T86 alloy were prepared by friction stir welding process under four different environments, i.e. conventional (C-FSW), liquid nitrogen (LN-FSW), ultrasonic vibration welding (U-FSW), and ultrasonic vibration with liquid nitrogen (ULN-FSW). The welding tool rotational speed of 650 rpm and tool traverse speed of 80 mm/min were fixed for all experiments. The microstructural characterization of the welded samples is examined with optical microscopy, SEM, and XRD. Higher tensile strength and microhardness were obtained in the ULW-FSW welded samples. The ultrasonic vibrations promote plastic deformation and material mixing, increasing the dislocation density and improving the weld strength. In addition, liquid nitrogen restricts the grain size and dissolution of strengthening precipitates due to the rapid welding cooling rate, resulting in higher hardness. Comparative analysis of weld joint efficiencies revealed that the ULN-FSW welded sample achieved an impressive efficiency of 87.12%. In contrast, U-FSW, LN-FSW, and C-FSW displayed efficiencies of about 86.17%, 83.01%, and 74%, respectively.

Keywords: Al-Cu-Li alloy; FSW; Vibration Welding; Liquid Nitrogen; Tensile Properties

1. Introduction

Al–Li alloy, with excellent mechanical properties and corrosion resistance, makes it suitable for the aerospace and automobile industries. The welding of Al–Li alloys by fusion welding methods causes various defects like porosity in the weld region, distortion, weld cracking, and loss of lithium due to evaporation [1, 2]. All the above defects are minimized by a solid-state welding technique, i.e. the FSW process [3, 4]. AA2099 is currently used for high-performance applications, such as the floor structure of Airbus' A380 aircraft because of its superior damage tolerance, high strength, and excellent corrosion resistance [5]. Naumov et al. [6] reported that the increased transverse speed of the tool increases the FSWed AA2099 weld joint strength due to a decrease in the sharp angle pattern of the remnant oxide line in the stir region. Lin and Zheng [7] FSWed the AA2099 alloy samples prepared by the liquid metallurgy route. The base alloy grain size was around 13 μm and was revealed in pancake grains. The 5 μm grain size occurs in the stir zone due to the stirring of elongated grains. Sidhar and Mishra [8] found that the underwater FSW of the Al–Cu–Li alloys HAZ and TMAZ was narrower than conventional FSW due to the cooling effect of water. The AA2195-T8 FSW joint failure occurs in the HAZ region due to the coarsening of Al2CuLi [9]. Gibson et al. [10] found that friction stir lap joints of Al–Li alloys welded with tapered tool shoulders have higher strength than flat shoulder joints. Sinhmar and Dwivedi [11] investigated the AA2014 alloy with a stationary FSW shoulder tool and found a superior weld strength joint than a conventional FSW tool shoulder. Less welding material heating occurred during welding due to non-flash formation and non-removal of plasticized material. Milagre et al. [12] welded AA2098 by the FSW process and found a density reduction of Al_2CuLi particles in the HAZ region, which decreased microhardness. Tao et al. [13] welded AA2060 by the FSW process and reported that the rotational speed could directly affect the tensile properties of the joint. Kumar et al. [14] welded AA2050 by the FSW process and concluded that the stir zone grain size decreases with increased tool traverse speed during welding. The heat input is also reduced without flash, resulting in a good joint with a good weld bead appearance. Ma et al. [15] studied the corrosion behaviour of the AA2099 plate joined by FSW and concluded that shallow pits were noticed in the

*Corresponding author, E-mail: yuvraj@dce.ac.in

Published online: 17 February 2024

Table 1 Chemical composition of the base material

Element	Li	Cu	Zn	Mn	Mg	Zr	Ti	Fe	Al
2099-T86	1.75	2.96	0.87	0.34	0.27	0.06	0.02	0.03	Rem

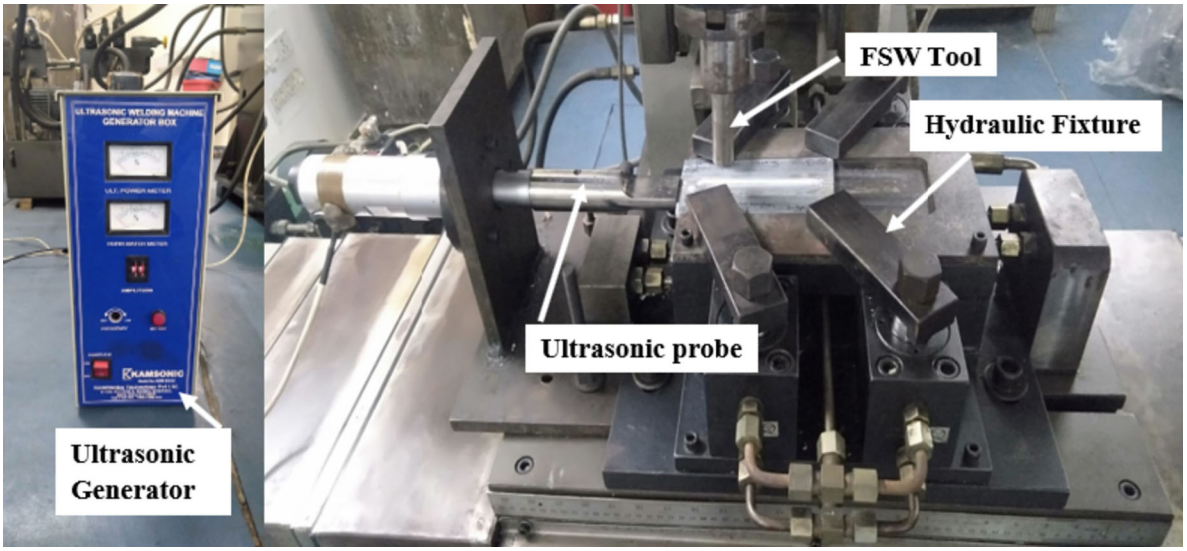


Fig. 1 Experimental setup

boundary of the stir zone and TMAZ due to the dissolution of the A_2CuLi phase. Cisko et al. [16] welded AA2099 by the FSW process and reported that the material expulsion from the nugget zone is due to high heat input. Sun et al. [17] worked on deep cryogenic treatment on AA2198 welded by FSW and concluded that deep cryogenic treatment could increase the weld strength, owing to a reduction in the grain size of the nugget zone. Snyder and Strauss [18] reported that water cooling during the FSW process could increase the strength and hardness of the FSW joints due to a decrease in heat generation in the weld region. The FSW plates were welded with the underwater condition, and the strength of the joint was improved due to less coarsening of the precipitates and reduced heat generation during water cooling [19–23]. Zhang et al. [24] welded AA2219Al-T6 by the FSW process under water cooling conditions and suggested that post-weld heat treatment produces more than 90% weld joint efficiency. Niu et al. [25] welded the AA2060 alloy by the FSW process and reported that the forced cooling FSW joint provides better tensile strength than the conventional FSW joint due to the grain size reduction in the nugget zone. Mofid et al. [26] worked on the friction stir welding of AA5083-H34 under different welding environments. They concluded that a lower peak temperature in the case of ultrasonic-assisted FSW with liquid nitrogen overpowers the formation of

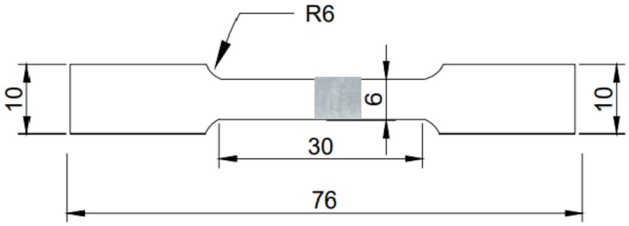


Fig. 2 Tensile test specimen dimensions

intermetallic compounds. Singh et al. [27] investigated the fatigue crack growth of the AA5052-H32 alloy during FSW with liquid nitrogen, water, and air cooling. The crack growth is higher in liquid nitrogen due to entrapped oxides in the nugget area. Bagheri et al. [28] observed a reduction in grain size when water and vibration were applied simultaneously during FSW welding of AA5083 alloy. Using ultrasonic-assisted brazing, Jie et al. [29] investigated the reinforcing impact of Mg alloy intermetallic compounds on Al alloy joints with Al/Zn/Cu/Zn multi-interlayer. The quick cooling caused by ultrasonic cavitation decreases the braze grain size and distributes the uniform mixing of the reinforcing particles. Similar results were obtained for friction stir vibration brazing of low-carbon steel using SiO_2 nano-reinforcing particles [30]. The mechanical vibration-assisted friction stir brazing

Table 2 Welding test results

Welded specimen/sample	Welding environments	UTS (MPa)	Grain size (μm)	NZ Microhardness (HV)	Joint Efficiency (%)
C-FSW	Air	236	51.63	124.3	74.00
LN-FSW	Ultrasonic vibration	263	47.18	145.5	83.01
U-FSW	Liquid nitrogen	273	41.55	149.5	86.17
ULN-FSW	Liq. N_2 with ultrasonic vibration	276	36.02	161.7	87.12

reduces the grain size of the low-carbon steel brazed joints due to increased heat input [31]. Zhao et al. [32] found that the tensile strength of the AA6063 alloy joints prepared by the ultrasonic-assisted liquid phase diffusion bonding method is equal to the strength of the parent metal. Bagheri et al. [33] examined AA6061-T6 welded joints fabricated through friction stir vibration welding (FSVW) and FSW techniques. The vibration frequency refines the grain size in the stir zone of FSVW samples, enhancing the welded joint's tensile strength and hardness. The mechanism behind this phenomenon was attributed to the intensified role of dynamic recrystallization in FSVW. Zhang et al. [34] examined the acoustic effects on heat transfer and material flow of FSW of dissimilar Al/Mg alloys. The acoustic effect increases the strength of the welded joint. Liu et al. [35] applied liquid CO_2 during FSW of AA1050-H24 alloy and found that cooling enhances the tensile strength of the welded joint and reduces their elongation due to the shrinks of the welded region. FSW is considered an alternative welding process for welding high-strength Al–Li alloys due to the absence of melting of base material. Various researchers have reported FSW of Al–Li alloys with multiple parameters for studies on enhancing weld joint efficiency. The environmental conditions during FSW significantly influence the mechanical properties of the welded joint. Limited work has been studied on the FSW of Al–Li alloys with combined cryogenic cooling media and vibrations. In this work, an attempt has been made to investigate the effect of liquid nitrogen cooling media and ultrasonic vibrations on the mechanical properties of the FSWed Al–Li alloy. This paper has attempted to compare the microstructural, hardness, and tensile strength of AA2099-T86 alloys welded with the FSW process in different environmental conditions and base materials.

2. Material and Methods

AA2099-T86, with a 6 mm thickness, was selected as a work material. The chemical composition of the base alloy is exhibited in Table 1. FSW machine capacity of 40 kN was used to fabricate the welded joints. The experimental setup is shown in Fig. 1. A H13 hardened threaded tool with a shoulder diameter of 20 mm and a pin diameter of 6 mm was used to join the samples. Before welding, the plates were cleaned with acetone. After many trial experiments, the rotational tool speed of 650 rpm and tool traverse speed of 80 mm/min were fixed. A standard metallographic procedure was adopted for the cross section of the welded samples for microstructural examination. The polished samples were etched with an etchant of 2 ml HF, 3 ml HCL, 20 ml HNO_3, and 175 ml water [36]. The etched samples were investigated with optical and scanning electron microscopes. The microhardness value of the cross section of welded specimens was measured with Vickers microhardness tester using 300 g of load with a dwell time of 20 s. The tensile specimens are extracted from the welded region are as per the ASTM-E8M standard [37].

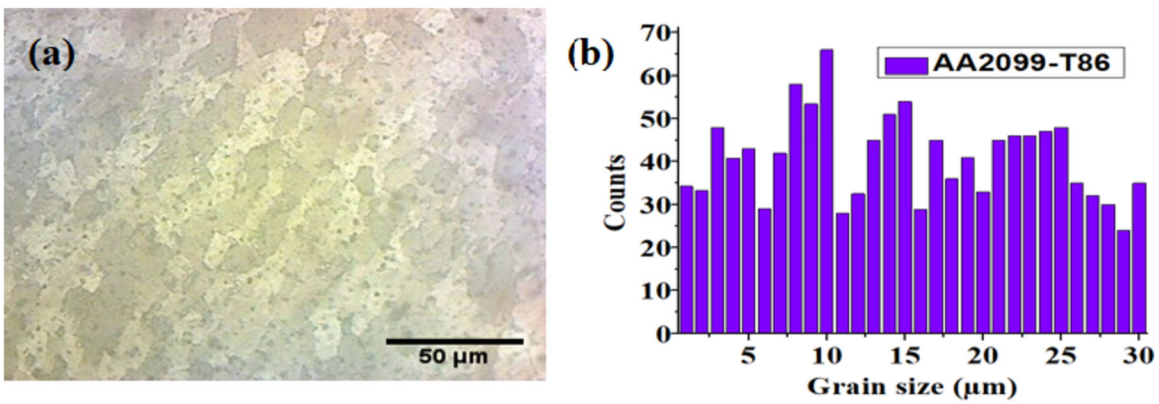


Fig. 3 **a** Microstructure of the AA2099-T86 base alloy and **b** grain size

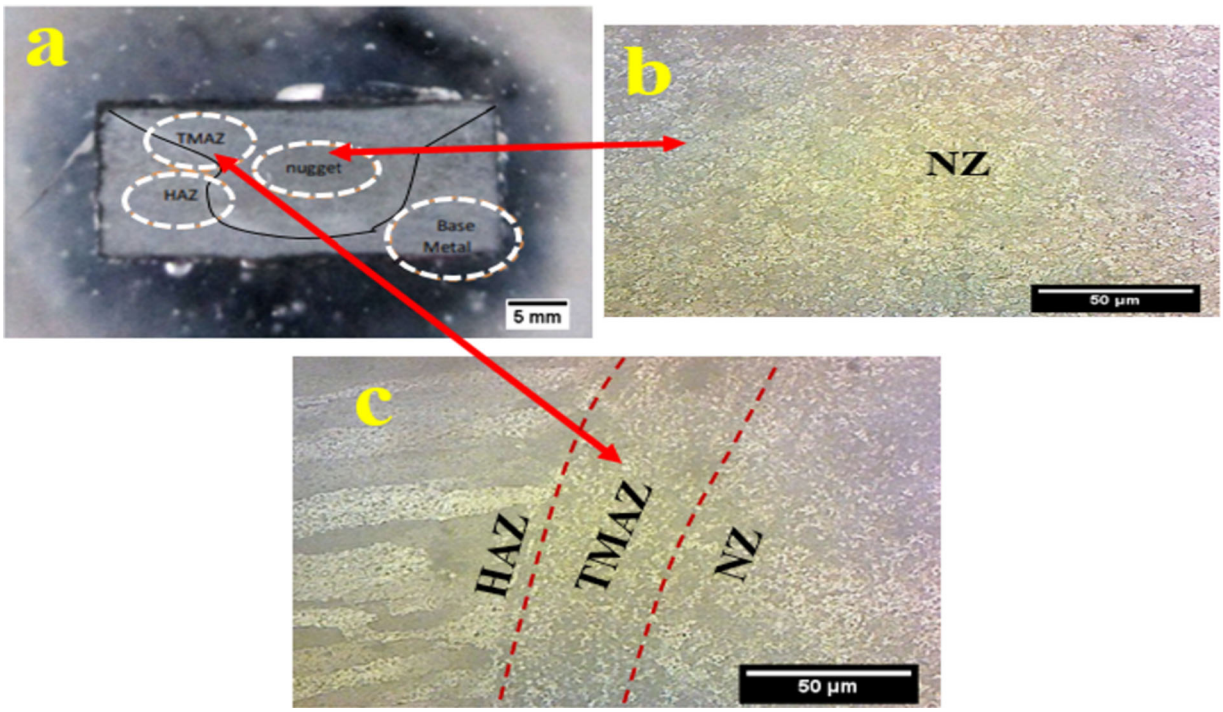


Fig. 4 Micrograph of FSW specimen (**a**) Macro-image (**b**) NZ (**c**) Interface region

The tensile test specimen dimensions are given in Fig. 2. The tensile tests of welded samples were conducted using a universal testing machine with a capacity of 50 kN at a 1 mm/min speed.

3. Results and Discussions

3.1. Microstructure Evaluation

FSW is used to produce fine, ultra-fine, and nanocrystalline refined materials. FSWed NZ microstructures are typically higher refined when compared to base material. The base alloy microstructural image is shown in Fig. 3. The grain size of the base alloy is around 67 µm. The macro- and microstructure of the welded samples taken at various zones are shown in Fig. 4. Macro-structure displays the distinct nugget zone with elongated grains of the TMAZ, which is formed due to induced strain from the stirring action of the tool pin during welding. The tool shoulder creates forging pressure, generating heat due to material melting. The rotating tool pin stirs the weld material, and mixing results in the formation of the joint. The material flow behind the FSW tool pin occurs during tool advancement. The stirring action of the material leads to dynamic recrystallization caused by severe plastic deformation, which results in refined equiaxed grains in the stir region [38, 39]. Figure 5 shows the nugget zone microstructures of joints fabricated by C-FSW, LN-FSW, U-FSW and ULN-FSW methods. Table 2 reveals that ULN-FSW joints exhibit the smallest grains, whereas C-FSW joints display the largest ones. The size of recrystallized grains is intricately linked to factors such as heat input, strain rate, and cooling rate. The temperature evolution during FSW welding also affects the microstructure characterization of the welded joints. The nugget zone is fully recrystallized, TMAZ partially recrystallized, and in HAZ, there is no recrystallization, but only grain coarsening occurs [40]. In the case of C-FSW, recrystallized grain growth occurred due to the elevated heat input and lower cooling rate providing ample heat input and time [41]. With the application of liquid nitrogen during FSW, the grain size of the stir region is reduced to 47.18 µm. The decrease in grain size may be due to the effective cooling of liquid nitrogen, which restricted the grains to the finer grains [42–44]. Further, the grain size of the stir region was reduced with ultrasonic vibration-assisted welding. The volume of the pin-driven zone is higher in U-FSW weld specimens due to the stress superposition and acoustic softening brought about by ultrasonic vibration during welding [45]. Incorporating vibration induces an intensified plastic material flow, improving weld quality. This enhancement is attributed to the augmented plastic material flow proximal to the welding tool [46]. In addition, temperature occurs during welding increases in material flow rate and efficient cooling provided by the

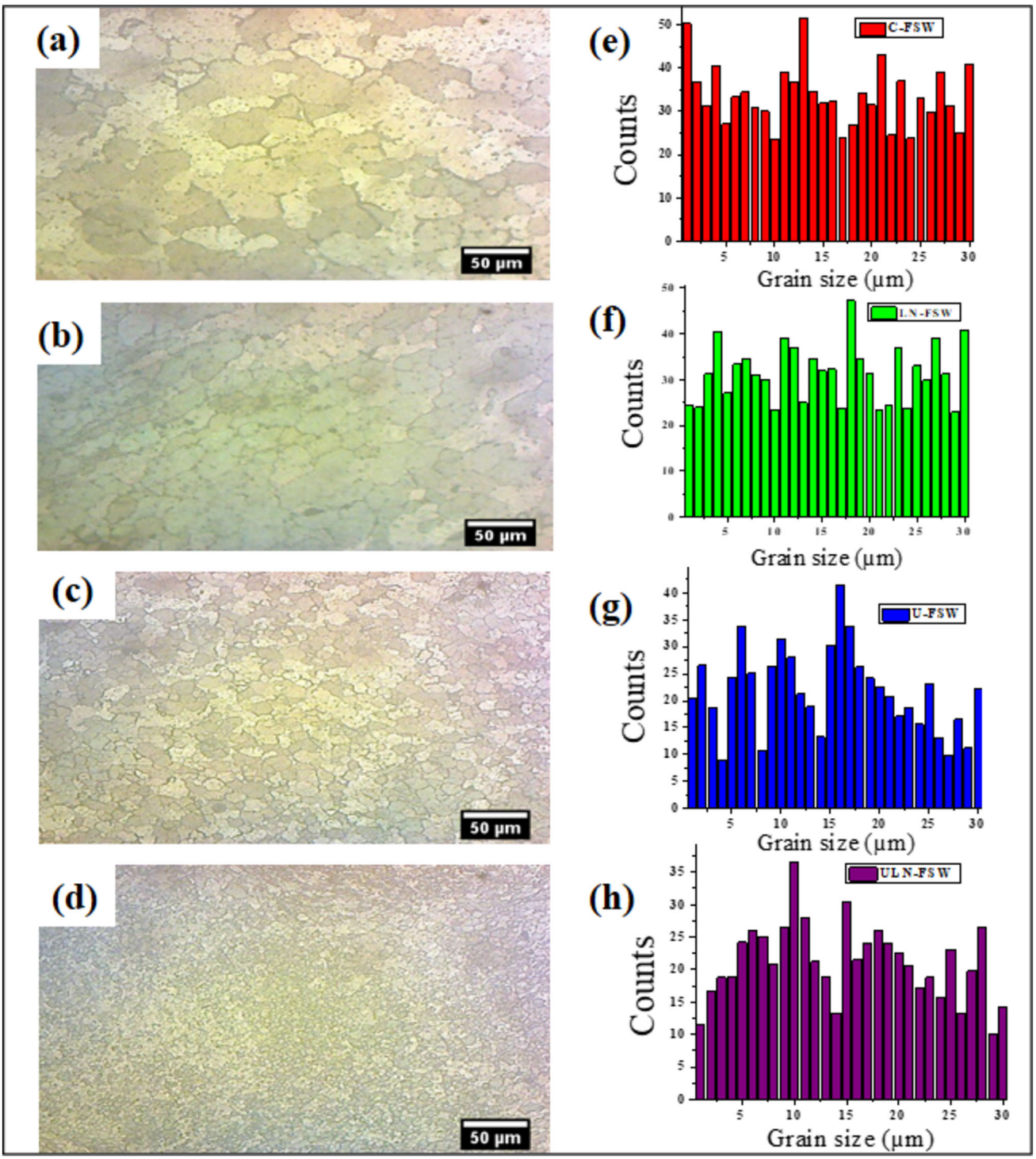


Fig. 5 Microstructure of the nugget zone (**a**) C-FSW (**b**) LN-FSW (**C**) U-FSW (**d**) ULN-FSW samples and (**e–h**) grain sizes of nugget zones

liquid nitrogen; the stir zone exhibits crooked morphology with ultra-fine grains when FSW is utilized together with vibration welding. The microstructure image revealed that the mixing of material is more uniform in the stir zone, which leads to uniformly recrystallized grains. Refined grain structures arise from elevated levels of dynamic recrystallization and increased plastic strain within the material during welding [47].

A lazy S type of welded defect is formed during the stirring of the material, which includes a dark zig-zag pattern near the root of the weld [48]. FESEM images show the presence of precipitates in the stir region of welded

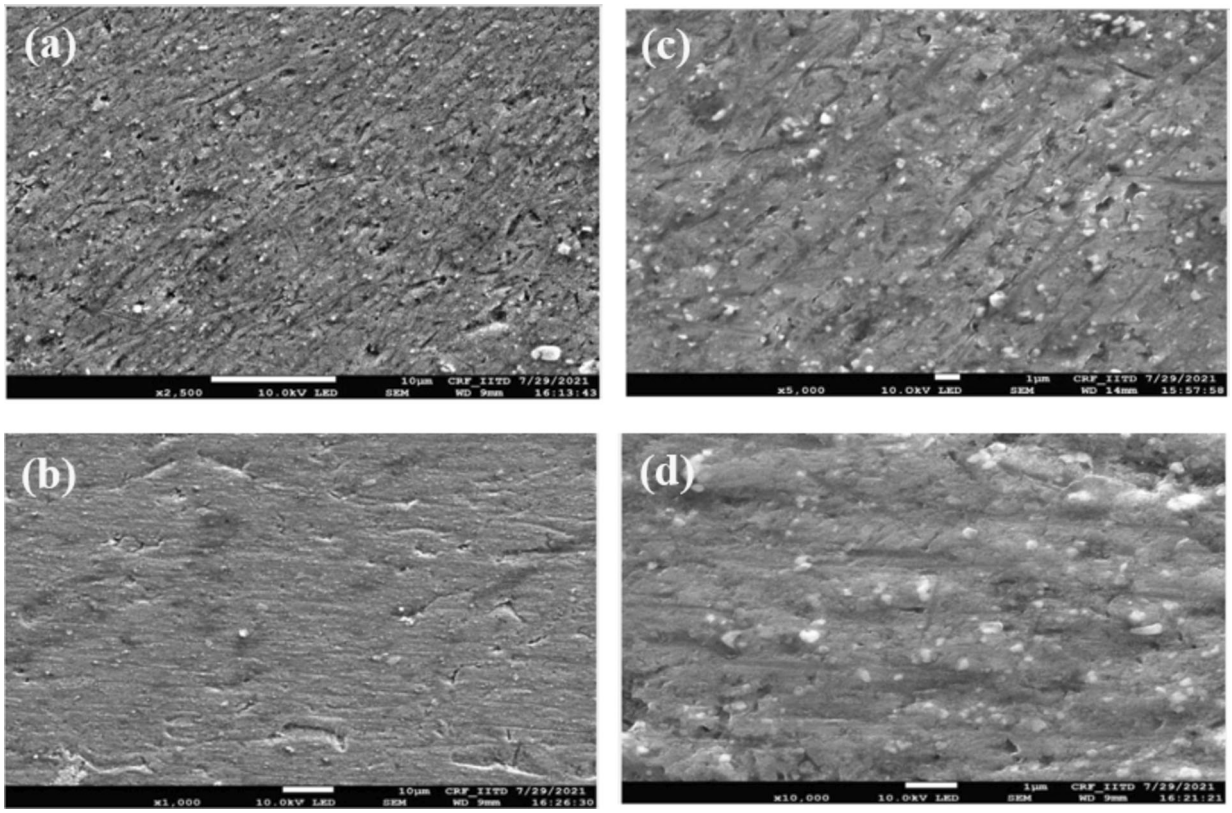


Fig. 6 FESEM images of nugget zone (**a**) C-FSW sample and **b** ULN-FSW sample; **c** and **d** are the enlarged images of Fig. **a** and **b**, respectively

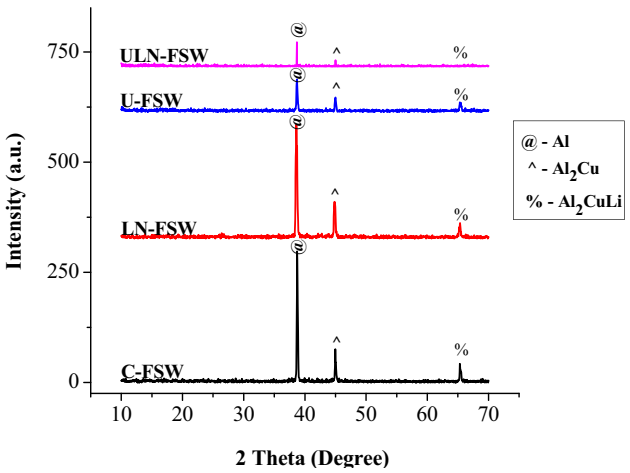


Fig. 7 XRD analysis of welded joint

samples, as seen in Fig. 6. Similar results were reported by Urminsky et al. [49] during the joining of AA2099 plates. In ULN-FSW-welded samples, finer precipitates are formed compared to other samples, decreasing grain size. The cross-sectional analysis of the welded samples involved X-ray diffraction (XRD) to discern precipitate formations under distinct environmental conditions. The XRD results revealed that Al_2Cu and Al_2CuLi precipitate in the stir zone (Fig. 7). Intermetallic compounds are formed at the interface of FSW welded joint due to the increased heat input at the stir zone of welded sample [50, 51].

3.2. Tensile Properties

Tensile-tested specimens before and after welding are shown in Fig. 8. The specimens' tensile test results are exhibited in Table 2. The stress–strain curve of tensile-tested specimens is depicted in Fig. 9(a). The tensile strength of all welded specimens is less than that of the base alloy due to precipitate ripening and annealing caused by the welding. The maximum tensile strength was 276 MPa in the ULN-FSW sample, with a joint efficiency of 87.12%. Due to fine grain formation, enhanced dynamic recrystallization, and less constitutional liquation, applying ultrasonic vibration and liquid nitrogen during FSW produced the maximum strength. Using liquid nitrogen restricts the grain size and dissolution of strengthening precipitates during welding, resulting in improved joint strength. The use of ultrasonic vibrations promotes plastic deformation and material mixing, which increases the dislocation density and results in improved strength. As per

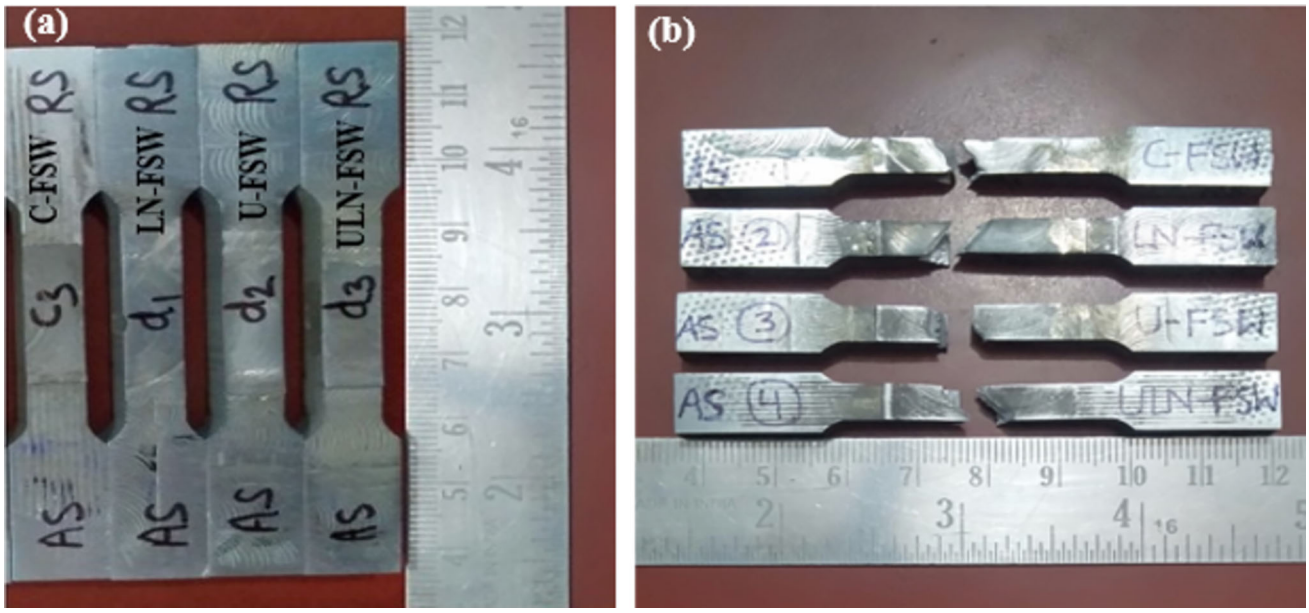


Fig. 8 Tensile test specimens (**a**) Before and **b** After testing

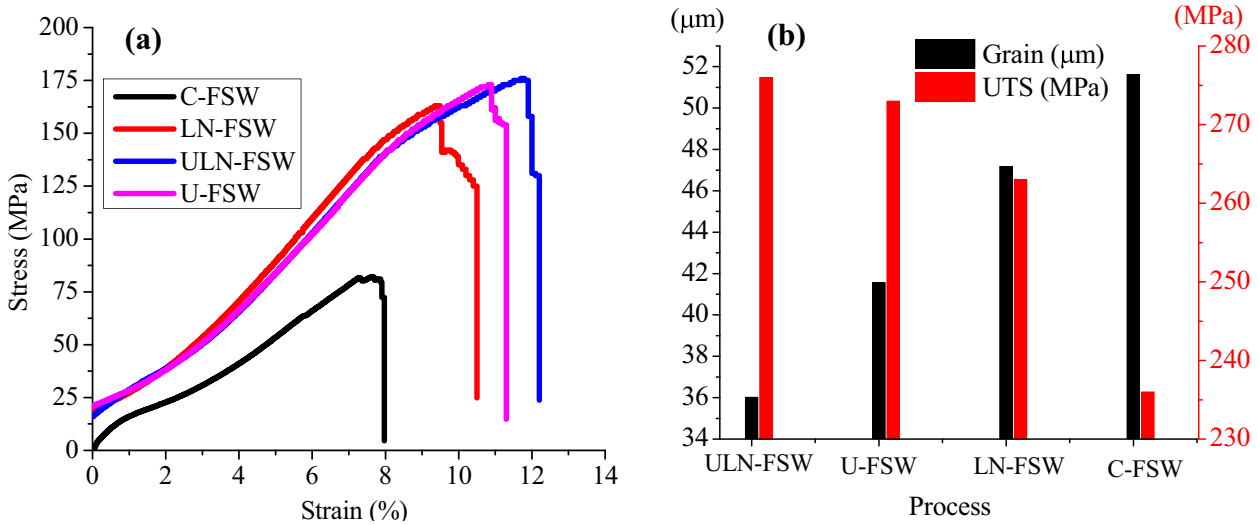


Fig. 9 **a** Stress–strain curve **b** Tensile strength and grain size graph

the Hall–Petch relationship, lower-grain welded samples have higher strength [52]. An analogous enhancement in tensile strength was observed in AA5083 welding through friction stir vibration welding with water cooling [46].

Figure 9(b) shows the relationship between the grain size of the stir region and the tensile strength of welded samples. It shows that lower-grain welded samples have higher strength. Without vibration-assisted liquid nitrogen cooling, welded sample strength has been slightly reduced to 83.01% joint efficiency. A lower tensile strength of 236 MPa was observed in the C-FSW sample. The maximum elongation of 12.5% was observed in the ULN-FSW sample, which is higher than other welded samples. Abdollahzadeh et al. [41] demonstrated similar findings in their investigation of AA6061-T6 alloy through friction stir welding under varying conditions. The outcomes indicated that the joint efficiency of the underwater friction stir welding configuration reached approximately 89%, surpassing 82% and 77% observed for friction stir vibration welding and conventional friction stir welding joints, respectively.

3.3. Microhardness

The microhardness of the welded samples taken at various regions is shown in Fig. 10. An increase in microhardness

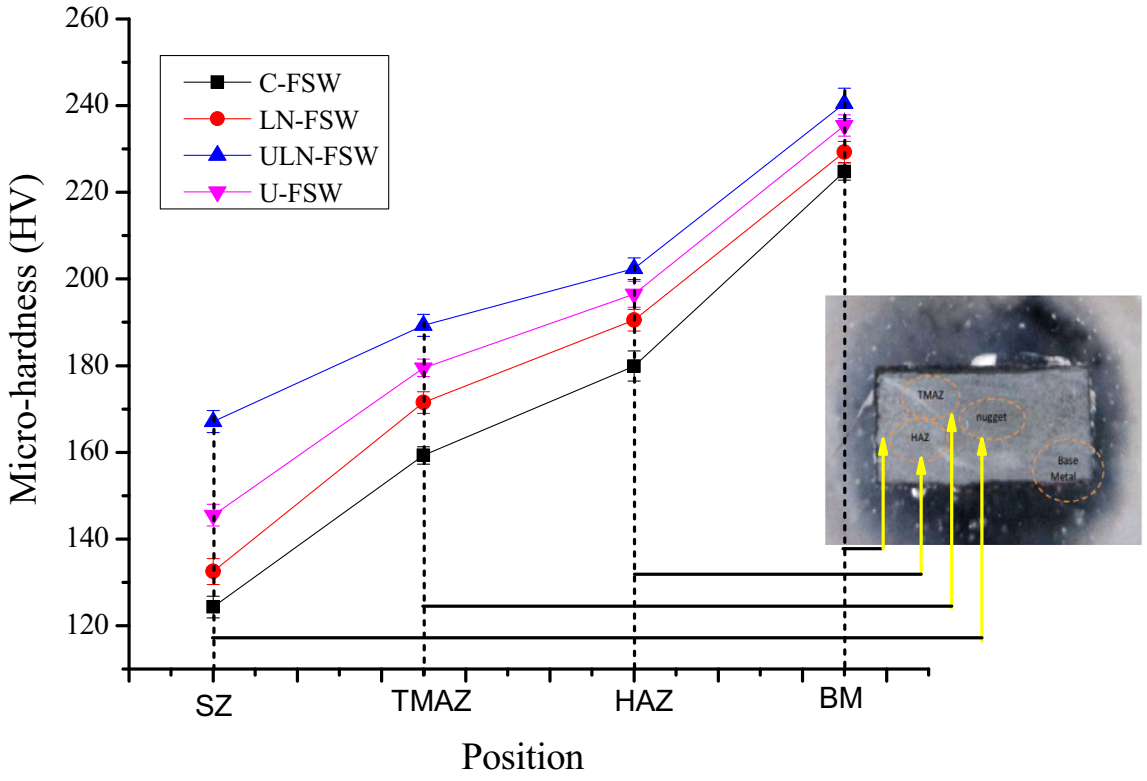


Fig. 10 Microhardness distribution of welded samples taken at various zones

was noticed in the order of NZ, TMAZ, HAZ, and the base metal of all welded samples. The material recrystallizing most likely causes lower microhardness in the stir zone due to precipitate coarsening. The dislocation movement decreased in the TMAZ and HAZ regions, and grain growth occurred in boundaries due to slow cooling, which increased hardness [38]. Conventional welding samples obtained lower microhardness value than other welded samples. The hardness value found in case of the U-FSW sample was greater than the LN-FSW and C-FSW due to the formation of fine precipitate of Al_2Cu and Al_2CuLi. With the liquid nitrogen cooling effect, the hardness value of welded samples was improved. The microhardness deviation also depends on the type of quenching medium and the metallurgical properties of the materials. A higher microhardness value was observed in the ULN-FSW sample. The ultrasonic vibration increases the material flow rate during welding and may cause the molten weld pool dislocations. In addition, cryogenic cooling rapidly cools the weld material, improving its hardness [36]. A comparable enhancement in microhardness values emerged from applying friction stir vibration welding with water cooling when welding of Al alloy [53]. Higher microhardness value achieved in this case, attributable to the synergistic effects of lower heat input, an adequate extent of precipitations, and a concurrently reduced grain size [54].

4. Conclusions

This study delved into the investigation of weld strength in AA2099-T86 alloy, employing the FSW process across four diverse environments. The findings gleaned from the experimental analyses shed light on crucial aspects:

(1) Comparative analysis of joint efficiencies revealed that the ULN-FSW welded sample achieved an impressive efficiency of approximately 87.12%. In contrast, U-FSW, LN-FSW, and C-FSW displayed efficiencies of about 86.17%, 83.01%, and 74%, respectively. Industrial welding applications can easily be scaled up using the easy and highly effective ULN-FSW welding method.

(2) The nugget zone's microstructure exhibited finer grains in the ULN-FSW sample, distinguishing it from the remaining welded samples.

(3) The ULN-FSW sample showcased a superior tensile strength of 276 MPa due to smaller precipitate size, ultra-fine grain formation in the stir region, and dynamic recrystallization. It reduced constitutional liquation, collectively contributing to enhanced joint strength.

(4) Microhardness assessments demonstrated a notable increase of 36% in the ULN-FSW sample and 27% in the LN-FSW sample compared to C-FSW. This

variance was attributed to the distinct development of intermetallic compounds.

References

[1] M.P. Alam and A.N. Sinha, Fabrication of third generation Al–Li alloy by friction stir welding: a review. Sadhana., *44* (2019) 1–13.

[2] N. Nayan, M. Yadava, R. Sarkar, S.N. Murty, N.P. Gurao, S. Mahesh and I. Samajdar, Microstructure and tensile response of friction stir welded Al–Cu–Li (AA2198-T8) alloy. Mater. Char., *159* (2020) 110002.

[3] B. Das, S. Pal and S. Bag, Probing defects in friction stir welding process using temperature profile. Sadhana, *44* (2019) 79.

[4] M.M. Moradi, H. Jamshidi Aval and R. Jamaati, Effect of tool pin geometry and weld pass number on microstructural, natural aging and mechanical behaviour of SiC-incorporated dissimilar friction-stir-welded aluminium alloys. Sadhana, *44* (2019) 9.

[5] J. Goebel, T. Ghidini and A.J. Graham, Stress-corrosion cracking characterization of the advanced aerospace Al–Li 2099–T86 alloy. Mater. Sci. Eng. A, *673* (2016) 16–23.

[6] A. Naumov, F. Isupov, E. Rylkov, P. Polyakov, M. Panteleev, A. Skupov, S.T. Amancio-Filho and O. Panchenko, Microstructural evolution and mechanical performance of Al–Cu–Li alloy joined by friction stir welding. J. Mater. Res. Tech., *9*(6) (2020) 14454–14466.

[7] Y. Lin and Z. Zheng, Microstructural evolution of 2099 AlLi alloy during friction stir welding process. Mater. Char., *123* (2017) 307–314.

[8] H. Sidhar and R.S. Mishra, Aging kinetics of friction stir welded Al–Cu–Li–Mg–Ag and Al–Cu–Li–Mg alloys. Mater. Des., *110* (2016) 60–71.

[9] C. Peng, C. Jing, Q. Siyi, Z. Siqi, S. Shoubo, J. Ting, Z. Zhiqing, J. Zhihong and L. Qing, Friction stir welding joints of 2195–T8 Al–Li alloys: correlation of temperature evolution, microstructure and mechanical properties. Mater. Sci. Eng. A, *823* (2021) 141501.

[10] B.T. Gibson, M.C. Ballun, G.E. Cook and A.M. Strauss, Friction stir lap joining of 2198 aluminum–lithium alloy with weaving and pulsing variants. J. Manuf. Proc., *18* (2015) 12–22.

[11] S. Sinhmar and D.K. Dwivedi, Mechanical behavior of FSW joint welded by a novel designed stationary shoulder tool. J. Mater. Proc. Tech., *277* (2020) 116482.

[12] M.X. Milagre, N.V. Mogili, U. Donatus, R.A. Giorjao, M. Terada, J.V. Araujo, C.S. Machado and I. Costa, On the microstructure characterization of the AA2098–T351 alloy welded by FSW. Mater. Char., *140* (2018) 233–246.

[13] Y. Tao, Z. Zhang, B.H. Yu, P. Xue, D.R. Ni, B.L. Xiao and Z.Y. Ma, Friction stir welding of 2060–T8 AlLi alloy. Part I: microstructure evolution mechanism and mechanical properties. Mater. Char., *168* (2020) 110524.

[14] S. Kumar, U. Acharya, D. Sethi, T. Medhi, B.S. Roy and S.C. Saha, Effect of traverse speed on microstructure and mechanical properties of friction-stir-welded third-generation Al–Li alloy. J. Braz. Soc. Mech. Sci. Eng., *42* (2020) 1–3.

[15] Y.L. Ma, H.B. Xu, Z.Y. Liang and L. Liu, Corrosion resistance of friction stir welded Al–Cu–Li alloy AA2099-T8. Acta. Metal. Sin., *33* (2020) 127–134.

[16] A.R. Cisko, J.B. Jordon, D.Z. Avery, T. Liu, L.N. Brewer, P.G. Allison, R.L. Carino, Y. Hammi, T.W. Rushing and L. Garcia, Experiments and modeling of fatigue behavior of friction stir welded aluminum lithium alloy. Metals, *9*(3) (2019) 293.

[17] J.Q. Sun, Y. Ma, C. Gao and H.Y. Luo (2019) Comprehensive tensile properties improved by deep cryogenic treatment prior to aging in friction-stir-welded 2198 Al–Li alloy. Rare. Met. 1–7.

[18] B. Snyder and A.M. Strauss, In-process cooling of friction stir extruded joints for increased weld performance via compressed air, water, granulated dry ice, and liquid nitrogen. J. Manuf. Proc., *68* (2021) 1004–1017.

[19] R. Saha and P. Biswas, Current status and development of external energy-assisted friction stir welding processes: a review. Weld. World., *66*(3) (2022) 577–609.

[20] H. Abidi, S. Konovalov and X. Chen, The effects of in-process cooling during friction stir welding of 7475 aluminium alloy. Sains. Mal., *50*(9) (2021) 2743–2754.

[21] M.A. Wahid and A.N. Siddiquee, Review on underwater friction stir welding: a variant of friction stir welding with great potential of improving joint properties. Trans. Non. Ferro. Met. Soc. China., *28*(2) (2018) 193–219.

[22] J.V. Christy, A.H. Mourad, M.M. Sherif and B. Shivamurthy, Review of recent trends in friction stir welding process of aluminum alloys and aluminum metal matrix composites. Trans. Non. Ferro. Metal. Soc. China., *31*(11) (2021) 3281–3309.

[23] V.P. Singh, S.K. Patel and B. Kuriachen, Mechanical and microstructural properties evolutions of various alloys welded through cooling assisted friction-stir welding: a review. Intermetallic., *133* (2021) 107122.

[24] Z. Zhang, B.L. Xiao and Z.Y. Ma, Enhancing mechanical properties of friction stir welded 2219Al-T6 joints at high welding speed through water cooling and post-welding artificial ageing. Mater. Char., *106* (2015) 255–265.

[25] S. Niu, Y. Yue, D. Yan, Z. Ma and S. Ji, Forced cooling friction stir welding of 2060–T8 Al–Li alloy. J. Mater. Eng. Perf., *28* (2019) 5763–5771.

[26] M.A. Mofid, A. Abdollah-Zadeh, F.M. Ghaini and C.H. Gur, Submerged friction-stir welding (SFSW) underwater and under liquid nitrogen: an improved method to join Al alloys to Mg alloys. Metall. Mat. Trans. A, *43* (2012) 5106–5114.

[27] R.K. Singh, R. Prasad, S. Pandey and S.K. Sharma, Effect of cooling environment and welding speed on fatigue properties of friction stir welded Al–Mg–Cr alloy. Int. J. Fat., *127* (2019) 551–563.

[28] B. Bagheri, M. Abbasi and M. Dadaei, Effect of water cooling and vibration on the performances of friction-stir-welded AA5083 aluminum joints. Metall. Micro. Anal., *9* (2020) 33–46.

[29] L.I. Jie, S.H.I. Lei, L.I. Yi-nan, T.S. Zhou, Y.U. Zi-jing, H.C. Zhang, Z.L. Peng, S.W.M. Miao, J.G. Gao and S.U. Zhao-fang, Reinforcement effect of Mg–Al–Zn ternary IMC on MB8/2024 joints with Al/Zn/Cu/Zn multi-interlayer via ultrasonic-assisted brazing. Trans. Non. Metals Soc. of China, *33* (2023) 1015–1028.

[30] V.S. Rizi, M. Abbasi and B. Bagheri, Investigation on intermetallic compounds formation and effect of reinforcing particles during friction stir vibration welding. J Mater. Eng. Perf., *31* (2022) 3369–3381.

[31] B. Bagheri, M. Abbasi, F. Sharifi and A. Abdollahzadeh, Different attempt to improve friction stir brazing: effect of mechanical vibration and rotational speed. Met. Mater. Int., *28*(9) (2022) 2239–2251.

[32] P. Zhao, Z. Li, Y. Xia, Z. Xu, X. Leng and J. Yan, Ultrasonic-assisted liquid phase diffusion bonding of Al alloys under low temperature and pressure for high-temperature resistance applications. Mater Design., *225* (2023) 111544.

[33] B. Bagheri, M. Abbasi and M. Dadaei, Mechanical behavior and microstructure of AA6061-T6 joints made by friction stir vibration welding. J. Mater. Eng. Perf., *29* (2020) 1165–1175.

[34] G. Zhang, J. Gao and C. Wu, Revealing the acoustic effects on heat transfer and material flow in ultrasonic assisted friction stir welding of dissimilar Al/Mg alloys. J. Mater. Res. Tech., *26* (2023) 1882–1902.
[35] X.C. Liu, W.T. Li, Y.Q. Zhou, Y.Z. Li, X.J. Pei, Z.K. Shen and Q.H. Wang, Multiple effects of forced cooling on joint quality in coolant-assisted friction stir welding. J. Mater. Res. Tech., *25* (2023) 4264–4276.
[36] A. Devaraju and V. Kishan, Influence of cryogenic cooling (liquid nitrogen) on microstructure and mechanical properties of friction stir welded 2014–T6 aluminum alloy. Mater. Today. Pro., *5*(1) (2018) 1585–1590.
[37] R.G. Narayanan, Effect of tool plunge depth on joint formation and mechanical performance of friction stir forming joints made between AA 5052–H32 and AA 6061–T6 sheet metals. Trans. Non. Ferro. Met. Soc. China., *28*(4) (2018) 613–628.
[38] A.K. Lakshminarayanan and V. Balasubramanian, An assessment of microstructure, hardness, tensile and impact strength of friction stir welded ferritic stainless steel joints. Mater. Des., *31*(10) (2010) 4592–4600.
[39] A.K. Lakshminarayanan, V. Balasubramanian and K. Elangovan, Effect of welding processes on tensile properties of AA6061 aluminium alloy joints. The. Int. J. Adv. Manuf. Tech., *40* (2009) 286–296.
[40] P. Chen, S. Zou, J. Chen, S. Qin, Q. Yang, Z. Zhang, Z. Jia, L. Zhang, T. Jiang and Q. Liu, Effect of rotation speed on microstructure evolution and mechanical properties of nugget zone in 2195–T8 Al–Li alloy friction stir welding joints. Mater. Char., *176* (2021) 111079.
[41] A. Abdollahzadeh, B. Bagheri, M. Abassi, A.H. Kokabi and A.O. Moghaddam, Comparison of the weldability of AA6061-T6 joint under different friction stir welding conditions. J. Mater. Eng. Perform., *30* (2021) 1110–1127.
[42] H. Ji, Y. Deng, H. Xu, S. Lin, W. Wang and H. Dong, The mechanism of rotational and non-rotational shoulder affecting the microstructure and mechanical properties of Al–Mg–Si alloy friction stir welded joint. Mater. Des., *192* (2020) 108729.
[43] C.C. Sastry, P. Hariharan, M. Pradeep Kumar and M.A. Muthu Manickam, Experimental investigation on boring of HSLA ASTM A36 steel under dry, wet, and cryogenic environments. Mater Manuf. Process., *34*(12) (2019) 1352–1379.
[44] M.J. Reza-E-Rabby, M.J. Olszta, N.R. Overman, M. McDonnell and S.A. Whalen, Friction stir dovetailing of AA7099 to steel with AA6061 interlayer for reduced Zn embrittlement at dissimilar interface. J. Manuf. Process., *61* (2021) 25–34.
[45] M.M. Abd Elnabi, T.A. Osman and A. El Mokadem, Evaluation of the formation of intermetallic compounds at the intermixing lines and in the nugget of dissimilar steel/aluminum friction stir welds. J. Mat. Res. Tech., *9*(5) (2020) 10209–10222.
[46] B. Bagheri, A. Abdollahzadeh, M. Abbasi and A.H. Kokabi, Numerical analysis of vibration effect on friction stir welding by smoothed particle hydrodynamics (SPH). Int. J. Adv. Manuf., *110* (2020) 209–228.
[47] B. Bagheri, M. Abbasi, F. Sharifi and A. Abdollahzadeh, Investigation into novel multipass friction stir vibration brazing of carbon steels. J Manuf Process., *37*(8) (2022) 921–932.
[48] G. Han, J.R. Lee, Y. Noh and T.S. Jun, The role of cryogenic quenching on the mechanical properties of FSWed 6061–T6 aluminum alloy. Mater. Sci. Eng. A, *840* (2020) 142896.
[49] J. Urminsky, M. Maronek, J. Barta, M. Sahul, F. Jurina and M. Pasak, Electron beam welding of aluminium alloy AW2099. Mater. Sci. For., *994* (2020) 28–35.
[50] B. Bagheri, M. Alizadeh, S.E. Mirsalehi, A. Shamsipur and A. Abdollahzadeh, The effect of rotational speed and dwell time on Al/SiC/Cu composite made by friction stir spot welding. Weld J., *66*(11) (2022) 2333–2350.
[51] Z. Zhang, C. He, Y. Li, L. Yu, S. Zhao and X. Zhao, Effects of ultrasonic assisted friction stir welding on flow behavior, microstructure and mechanical properties of 7N01-T4 aluminum alloy joints. J. Mater. Sci. Tech., *43* (2020) 1–13.
[52] C. Du, X. Wang, Q. Pan, K. Xue, M. Ni and J. Liu, Correlation between microstructure and mechanical properties of 6061–T6 double-side FSW joint. J. Manuf. Process., *38* (2019) 122–134.
[53] B. Bagheri, M. Abbasi and A. Abdollahzadeh, Microstructure and mechanical characteristics of AA6061-T6 joints produced by friction stir welding, friction stir vibration welding and tungsten inert gas welding: a comparative study. Int. J. Min. Met. Mat., *28* (2021) 450–461.
[54] B. Bagheri, M. Abbasi, A. Abdollahzadeh and A.H. Kokabi, A comparative study between friction stir processing and friction stir vibration processing to develop magnesium surface nanocomposites. Int. J. Miner. Metall., *27* (2020) 1133–1146.

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

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

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/378156862

Microplastics influence the functional responses of a tropical estuarine calanoid Pseudodiaptomus annandalei OPEN ACCESS EDITED BY

Article · February 2024

DOI: 10.3389/fevo.2024.1277332

CITATIONS
0

READS
139

3 authors, including:



Jaigopal Sharma
Delhi Technological University
151 PUBLICATIONS **2,082** CITATIONS
SEE PROFILE



Malayaj Rai
Silliman University
1 PUBLICATION **0** CITATIONS
SEE PROFILE

All content following this page was uploaded by Malayaj Rai on 13 February 2024.

The user has requested enhancement of the downloaded file.

frontiers | Frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 05 February 2024
DOI 10.3389/fevo.2024.1277332

Check for updates

OPEN ACCESS

EDITED BY
Jiang-Shiou Hwang,
National Taiwan Ocean University, Taiwan

REVIEWED BY
Devesh Kumar Yadav,
Central University of South Bihar, India
Ahmet Erkan Kideys,
Middle East Technical University, Türkiye

*CORRESPONDENCE
Jaigopal Sharma
sharmajaigopal@dce.ac.in

†PRESENT ADDRESS
Malayaj Rai,
Department of Microbiology, University of Hawai'i at Mānoa, Honolulu, HI, United States

‡These authors have contributed equally to this work

RECEIVED 14 August 2023
ACCEPTED 08 January 2024
PUBLISHED 05 February 2024

CITATION
Sharma J, Rai M and Guino-o RS II (2024) Microplastics influence the functional responses of a tropical estuarine calanoid *Pseudodiaptomus annandalei*. *Front. Ecol. Evol.* 12:1277332. doi: 10.3389/fevo.2024.1277332

COPYRIGHT
© 2024 Sharma, Rai and Guino-o. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Microplastics influence the functional responses of a tropical estuarine calanoid *Pseudodiaptomus annandalei*

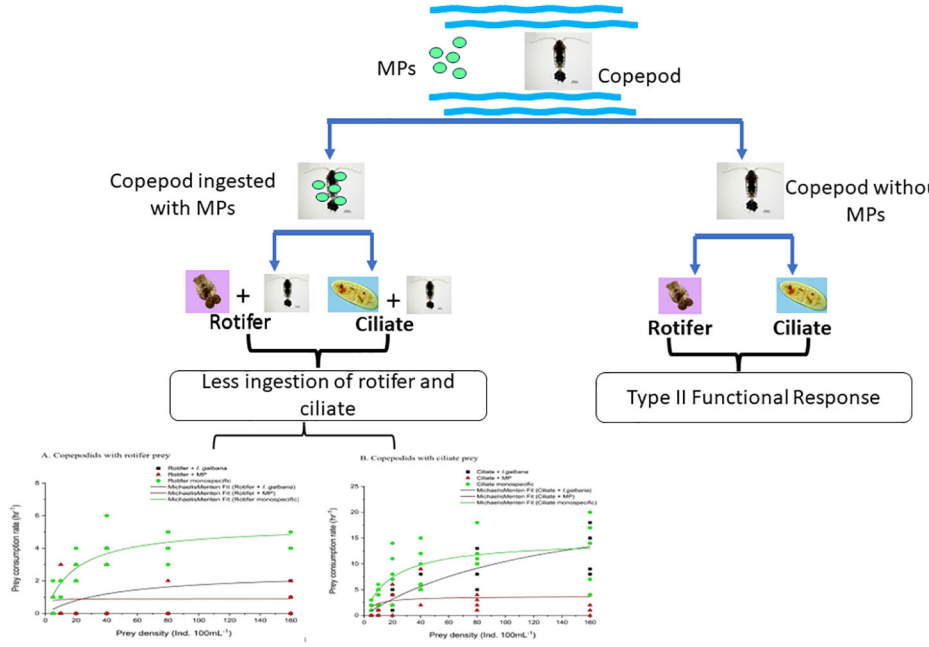
Jaigopal Sharma[1*‡], Malayaj Rai[2†‡] and Robert S. Guino-o II[2]

[1]Department of Biotechnology, Delhi Technological University, Delhi, India, [2]Biology Department, Silliman University, Dumaguete, Philippines

Being chemically inert and morphologically similar (<5 mm in size) to the natural prey of copepods, microplastics (MPs) affect them through entanglement, ingestion, and interference with their natural prey selection. The effects of external factors, such as MP, can be observed in the form of alterations in the functional and numerical responses of copepods. Functional response is explained as the relationship between the intake rate of a consumer and the amount of food available in a particular ecotope. Considered three types of functional response: (i) with increasing food density, when intake rate increases linearly until asymptote, called Type-I functional response; (ii) when the proportion of prey ingested by the consumer decreases monotonically with prey density, considered as a Type-II functional response; and (iii) when ingestion rate and food density show a sigmoid relationship, is considered as Type-III functional response. In the present study, we attempted to elucidate the effects of MP on the functional responses of brackish water, demersal, calanoid copepod, *Pseudodiaptomus annandalei*, feeding on the rotifer *Brachionus rotundiformis*, and the ciliate *Frontonia microstoma*. The ability of *P. annandalei* to survive environmental fluctuations and its nutritional value make it an ecologically important organism in estuarine ecosystems. The experiment for estimation of functional response included four setups with adult and copepodite stages of *P. annandalei*, each with six replicates at six different prey densities, in the presence and absence of MP particles. Without MPs, *P. annandalei* showed a typical Type-II functional response, where the proportion of prey consumed significantly declined with increasing prey density. In the MP-applied medium, the proportion of prey consumed did not exhibit a significant relationship with the initial prey density. The number of ciliates and rotifers ingested by copepodites and adults was significantly lower in MP-contaminated medium than in prey either monospecific or in combination with microalgae. Because of the comparable size of MP to the prey species and unspecialized feeding of *P. annandalei* in natural waters, we hypothesized reduced predation rates as a result of pseudo-satiation resulting from MP ingestion leading to malnutrition and MP storage in copepod biomass.

KEYWORDS

copepods, microplastics, functional response, foraging, ecosystem



GRAPHICAL ABSTRACT

Highlights

- Microplastic (MP) has negative effects on prey consumption rates of *Pseudodiaptomus annandalei*
- *P. annandalei* exhibited *type*-II functional response to monospecific prey and in the presence of microalgae
- MPs *per se* caused deviation from Type-II functional response
- MP contamination caused a ~20%–95% reduction in prey consumption by *P. annandalei* copepodites and adults.
- MP imposed alterations in the functional response type and prey ingestion can be used as sensitive endpoints to establish the impact of MP pollution.

Introduction

Microplastic (MP) particles have been reported from surface waters to the deepest trenches in every ocean (Auta et al., 2017; Richon et al., 2022). The rise in oceanic contamination by these small (<5 mm) and sometimes down to the micro- to nanometer scale (Barnes et al., 2009; Ter Halle et al., 2017) anthropogenic debris may lead to widespread contamination of oceanic food webs. MP particles or fibers have become emergent pllutants in estuarine and marine ecosystems. Food-web contamination by MP has been observed in fish and other higher trophic levels (Neves et al., 2015; Ugwu et al., 2021; Nair and Perumal, 2022; Avignon et al., 2023).

Recent studies have found MPs in wild-caught and commercial fish stomachs and guts (Jabeen et al., 2017; Baalkhuyur et al., 2018; Halstead et al., 2018; Kumari et al., 2023b), indicating large-scale MP contamination of the oceanic food web caused by direct MP ingestion (Roch et al., 2020), indirect MP ingestion, such as consumption of contaminated prey, or MP entanglements (Setälä et al., 2014; Nelms et al., 2018; Zhang et al., 2019). The ubiquitous nature of MP pollution in air, water, and soil has become a major concern among environmental activists, scientists, and citizens alike (Cunningham et al., 2020). In estuarine and brackish water ecosystems, copepods and rotifers play an important role in linking nano-and microplankton to higher trophic levels (Beaugrand et al., 2003; Tseng et al., 2011). Food-web contamination may begin with MP ingestion by zooplankton, which is the primary consumer and constitutes the lowest heterotrophic trophic level of aquatic ecosystems (Kumar, 2005; Amélineau et al., 2016; Richon et al., 2022). Zooplankton contamination with MPs may also have several biological implications, such as decreased growth, reproduction, and lower survival rates (Cole et al., 2019; Isinibilir et al., 2023). Their position in the food web is at the interface between the primary producers and higher trophic levels (Kumar, 2005; Kerfoot and DeMott, 2019). Thus, MP ingestion by zooplankton may act as an entry point for MPs into the marine food web (Gunaalan et al., 2023).

The ingestion of inert particles such as plastic beads by calanoid copepods has been frequently documented (Bundy and Vanderploeg, 2002; Ferrao Filho et al., 2005). With the help of mechanoreception ability, calanoid copepods can detect boundary layer disturbances and ingest non-motile inert particles in proximity to the first antennae of copepods (Bundy et al., 1998; Ayukai, 2020). These copepods redirect and ingest large inert particles entrained in their feeding currents (Vanderploeg et al., 1990). MPs are chemically inert and morphologically similar to the

natural food of copepods, and may affect them through entanglement, ingestion, and interference with other natural food organisms (Bhuyan, 2022). The smaller size and higher availability of MPs increases the probability of ingestion by marine organisms (Woods et al., 2020). The patchy heterogeneous distribution of MPs in estuaries (Akdogan et al., 2023; Rajan et al., 2023) and oceans (Pan et al., 2021) may also lead to zooplankton in some parts of the ocean being much more vulnerable to MPs than the rest. The extent of spread is also on the rise at an alarming rate, with more recent studies such as Kanhai et al. (2020) showing the presence of MPs even in the sea ice cores of the Arctic central basin. A range of zooplankton taxa have been shown to be able to ingest MPs of similar size to algal prey (7 µm–30 µm; 650 MP mL^{-1}–3,000 MP mL^{-1}), while smaller microplastics (3.8 µm; 40 MP $mL^{-1} \times 10^3$ MP mL^{-1}) can externally adhere to a copepod's functional appendages (Cole et al., 2015). Ingestion of plastic or latex beads, used as mimics or tracers of prey, by copepods has been reported by Frost (1977); Paffenhöfer and Van Sant (1985), and Cole et al. (2015). The feeding capacity and metabolic rates of calanoids, *Centropages typicus* and *Calanus helgolandicus* were significantly reduced in the presence of 6 µm–12 µm polystyrene beads (>4,000 MPs mL^{-1}) (Cole et al., 2013; Isinibilir et al., 2020). However, another recent study by Svetlichny et al. (2021) showed an increase in the consumption rate of pure microplastics in a mixture of microplastic beads (6µm diameter) and algae *Rhodomonas salina* (5 µm–10 µm size range) with equal concentrations of approximately 5,000 cells/beads mL^{-1}. The presence of micro- and nano-sized polystyrene in the environment has been observed to alter the feeding behavior of diving beetle (*Cybister japonicas*) and crucian carp (*Carassius carassius*) (Kim et al., 2018). The negative effects of MPs on physiology, organ damage, and even mortality have been shown in the laboratory (Franzellitti et al., 2019; Isinibilir et al., 2023), but only limited information is available on the behavioral impacts, feeding response, and functional and numerical responses of brackish water copepods commonly occurring in regions with higher MP pollution (Franzellitti et al., 2019; Cunningham et al., 2021; Rajan et al., 2023).

The ingestion of MP by zooplankton not only affects the aquatic food web, but also the entire ecosystem (Alava et al., 2023). Isinibilir et al. (2020) highlighted the negative effects of microplastic intake on the energy metabolism and swimming frequency of the calanoid copepod *C. helgolandicus*. As feeding is fundamental to the energetic requirements of an organism, MPs are likely to negatively affect zooplankton health by inducing a false sense of satiety, bioaccumulation in organisms, and blocking gut passage. Since copepods are an important component of the fish diet (Kerfoot and DeMott, 2019), they can also lead to trophic transfer of MPs to higher trophic levels, thereby posing a threat to the ecological functioning of the entire aquatic ecosystem.

Most feeding studies related to food selection have shown that copepods can distinguish between algal-flavored and unflavored artificial particles (DeMott, 1986), algae and polystyrene microspheres of similar size (DeMott, 1988), live and dead prey of the same species (Paffenhöfer and Van Sant, 1985; DeMott, 1988), toxic and non-toxic algae (Fulton and Paerl, 1987; DeMott and Moxter, 1991; Kumar, 2003), nitrogen-sufficient and nitrogen-limited algae of the same species (Cowles et al., 1998), algae and ciliates (Hartmann et al., 1993), ciliates and rotifers (Rao and Kumar, 2002), and larger and smaller rotifers (Dhanker et al., 2012). In contrast, some studies have reported that calanoid copepods are unable to exercise selection in regions with higher levels of suspended solids (Arendt et al., 2011) and tidally influenced estuaries with high sediment concentrations.

Previous studies have shown the effect of algae on animal prey consumption rate and alterations in the functional response of copepods (Kumar and Rao, 1999; Enríquez-García et al., 2013). Being numerically dominant and a preferred live food for fish larvae, *Psudodiaptomus annadalei* exerts a significant impact on the estuarine food web in the Indo-Pacific. Recent studies have demonstrated the predation efficiency of *P. annandalei* in natural setups, and ciliates and rotifers are important food items in their diet (Dhanker et al., 2012). *P. annandalei* is a generalist species with a large dietary niche breadth ranging from algal size, i.e., 7 µm–20 µm (Lee et al., 2010; Beyrend-Dur et al., 2011; Dhanker et al., 2015), to the size (150 µm–220 µm) of a rotifer (Dhanker et al., 2012). *P. annandalei* is able to select similarly sized rotifer over ciliate diet and show prey switching behavior on the basis of relative abundance of prey species. Dhanker et al. (2012) determined that the predatory success of *P. annandalei* on rotifers and ciliates followed an inverse size function, as observed for other invertebrate predators (Rao and Kumar, 2002; Hwang et al., 2009). However, there is limited knowledge regarding its ecology and functional feeding response. These organisms are currently facing potential threats from rapidly increasing MP concentrations in their immediate surroundings. MPs have been recorded in various domains of marine and estuarine environments (Li et al., 2019), which are supposed to be the habitat of the calanoid copepod *P. annandalei*, in the tropical and subtropical Indo-Pacific region. Another calanoid copepod, *C. helgolandicus*, showed a gradual decrease in food consumption rate during the 6-day experiment, feeding on a mixture of the microalgae *R. salina* and MP (Isinibilir et al., 2020). In a short-term experiment to evaluate the effect of initial MP concentration on consumption rate, Isinibilir et al. (2020) recorded a linear increase in MP consumption rates in *C. helgolandicus* with increasing concentration of MP in the medium, showing a Type-I functional response. Recent studies have shown the ability of copepods to distinguish between the chemical properties of MPs with the help of their mouthparts (Xu et al., 2022); however, despite the study showing the preference of zooplankton for its natural prey when both MP and natural prey are present, they do not shed light on the effect of MPs on the overall functional response of zooplankton. However, it is worth noting that the functional response can be best assessed in the presence of alternative prey (Iyer and Rao, 1996; Chan et al., 2017) as the functional response elucidates predator feeding rates with varying food densities (Solomon, 1949; Smith and Smith, 2020), thus, providing an estimation of energy transfer across trophic levels, understanding the stability and mechanism of coexistence of prey and predators (Oaten and Murdoch., 1975; Kumar and Rao, 2001). Xu et al. (2022) compares the ingestion of natural prey to MPs, it does not attempt to address and compare the number of natural prey consumed in presence and absence of MP.

Current knowledge of MP-imposed impacts on zooplankton largely focuses on the potential of zooplankton to engulf plastics. Most studies have used ingestion rates of MP as a parameter to describe the effects of MP on copepods (Jeong et al., 2017; Li et al., 2019).

Scientific knowledge of the functional response to a gradient of food density in MP-contaminated environment is essential to elucidate the adaptability of *P. annandalei* in an MP-contaminated environment and its response to the differential density of natural food. To date, no information is available on the estuarine calanoid, *P. annandalei*, whose natural habitats are contaminated by MP. A thorough understanding of the effects of MPs on the feeding ecology and foraging strategy of micro-meso-sized organisms is relevant to understand the effects of MPs on ecosystem functioning. To elucidate the impact of MPs on ecosystem functioning and feeding ecology, it is important to address the effects on the functional and numerical responses of an organism at the base of the food chain. Herein, we demonstrated the effects of MPs falling within the dietary niche size of the copepod *P. annandalei*. We also compared the effects of microalgae as an alternative food type in the medium to determine whether MP-imposed effects are similar to, or different from, alternative natural diets.

Thus, in the present study, we investigated the effects of MPs on the functional responses of *P. annandalei* (adult and copepodid stages, C2–C3) feeding on the rotifer *B. rotundiformis* and ciliate *F. microstoma*. We hypothesized that the presence of MPs along with the rotifer or ciliate would negatively influence prey ingestion rates in the copepod, *P. annandalei*, by inducing a false sense of satiety upon uptake of MPs, which would be reflected as an alteration in its functional response.

Materials and methods

Culture of experimental organisms

All experimental organisms were cultured in the laboratory three months prior to the start of the experiment. The starter cultures of the copepod *P. annandalei* (Sewell, 1919) and the rotifer *B. rotundiformis* (Tshugunoff 1921) were originally obtained from a coastal brackish water pond (salinity ~18 psu–20 psu) by filtering water through a 70 µm mesh plankton net. Field-collected animals were acclimatized to laboratory culture conditions. Pure cultures of the rotifer *B. rotundiformis* were performed in a 2-L beaker at 18 psu-20 psu salinity and 28°C ± 1.5°C using food from the microalgae *Isochrysis galbana* Parke (length: 5 µm–8 µm; width: 2.5 µm–4 µm). The copepod cultures were expanded to 20 L round cylindrical plastic tanks. A mixture of *I. galbana* (6–10 × 10^4 cells mL^{-1}) and rotifer *B. rotundiformis* was used as food for the copepods. The copepod culture was mildly aerated continuously to mix and maintain the food uniformly distributed in the culture tank. The culture medium was renewed with a mixture of filtered seawater (0.45 µm Millipore filters) and autoclaved freshwater twice a week. All cultures were maintained at ambient temperature (annual range: 20°C–30°C) under a 12L:12D. Mass cultures of the algae, *I. galbana*, were established in the laboratory. The algal culture was maintained in a 2 L borosilicate glass flask in Walne medium (Walne, 1970). Cultures of algae were maintained at 20–22 salinity at 25°C –32°C at 1,000 lx–20,000 lx under 12 h light:12 h dark photoperiodic conditions. The algae were harvested in the log phase of growth, sedimented by centrifugation, and resuspended in experimental medium. The ciliated protist *Frontonia microstoma* was continuously cultured at 20–22 salinity and 25°C using the algae *I. galbana* and adding Hay infusions (10%) to the culture medium. Table 1 lists the attributes of the grazer, prey type, and prey density used in the experiment.

MP preparation

Commercial polystyrene and polyethylene spheres were used in the experiment. The bead sizes (15 µm–30 µm) used in each experiment were selected to be comparable to the algal-food size range of *P. annandalei*. The encounter rate depends on the ratio of MP to zooplankton; therefore, our experimental protocol represented different ratios of ciliate and rotifer diets. In one set of treatments (i.e., prey ingestion in the presence of MP) with six replicates, MP particles were carefully counted and added to each treatment medium at a concentration of 100 particles mL^{-1} in each experimental vessel.

Experimental design

To understand the effect of MP on functional response of *P. annandalei* adult (females) and copepodites, following experiments were conducted (i) Feeding rates of copepodites (C3–C4 stage) on

TABLE 1 Taxonomic group, body length (Mean ± SE) and density of experimental organisms, and experimental protocol followed of the functional response experiment.

Group	Species	Body Length (µm)	Experimental protocol: No. of replicate-06 Density used (ind. 100 ml^{-1})					
Copepoda: Calanoida	*Pseudodiaptomus annandalei* (Adults)	1,128 ± 63 (n = 18)	3	3	3	3	3	3
Copepoda: Calanoida	*P. annandalei* (Copepodites)	643 ± 75 (n = 18)	3	3	3	3	3	3
Protista: Cilaiata	*Frontonia microstoma*	218 ± 39 (n = 20)	5	10	20	40	80	160
Rotifera: Monogononta	*Brachionus rotundiformis*	156 ± 18 (n = 20)	5	10	20	40	80	160

ciliate *F. microstoma* (ii) feeding rates of copepodites (C3–C4 stage) on rotifer *B. rotundiformis*, (iii) feeding rates of adult *P. annandalei* on ciliate *F. microstoma*, and (iv) feeding rates of adult *P. annandalei* on rotifer *B. rotundiformis*.

A batch of copepods (n = 180) of the same age group and previously exposed to the prey species to be used in the experiments were removed and kept in filtered water for 3 **h** for starvation and acclimation to the experimental volume to ensure uniform acclimation and hunger levels. Prior to the functional response experiment, the optimal feeding duration was standardized using a satiation time experiment. Based on the results obtained from the preliminary satiation time experiment, a 120-minute duration of feeding was followed for all the feeding trial experiments.

Functional response estimation

For both ciliate and rotifer prey types, six selected prey densities (Table 1) were offered to 3 h prestarved *P. annandalei* to elucidate the effects of MP on the functional response (response to changing prey density in the environment) of *P. annandalei*. Two sets of experiments were conducted in phases, consisting of (i) adult and (ii) copepodite stages of *P. annandalei*, each with three diet combinations (monospecific prey, presence of algae, and presence of MP), making a total of 108 (three prey combinations × six prey concentrations × six replicates) test jars of 150 ml capacity, containing 100 ml medium. For functional response studies involving copepodites, the experimental design was similar and copepodites at stages C3–C4 were used for the experiment. Prey concentrations set up in the geometric series (Table 1) were chosen based on preliminary experiments to cover the steeply rising and satiated portions of the functional response. All prey species were individually counted and introduced into the test jars at the desired density using a stereomicroscope. The experimental vessels were placed on black paper to provide contrast and minimize the prey stratification caused by uneven illumination. The copepods (three individuals per vessel) were then gently introduced into the experimental vessels and allowed to feed for 120 min. All the experimental vessels were covered with Perspex sheets, and light was provided by overhead fluorescent illumination. The number of prey eaten was estimated as the difference between the initial and final prey numbers after allowing copepods to feed for 120 min in the test jars. At the end of the experiment, copepods were anesthetized with 2-phenoxy ethanol and then preserved in buffered formalin for later gut content analysis and validation of the estimated number of prey eaten during the experimental period. The ciliate *F. microstoma* was kept in separate experimental vessels at the same density as that used for the experiment in quadruplicate to monitor any multiplication during the experimental period. Prey consumption by the predator was expressed as the number of prey items (rotifers and ciliates) consumed per 120 min.

Statistical analyses

The data on functional response were analyzed in two steps. First, to discriminate the type of functional response (Holling, 1959), the positive or negative sign of the linear coefficient was determined by logistic regression of the proportion of prey consumed (N_a/N_t) as a function of prey density (N_t):

$$\frac{N_a}{N_t} = \frac{\exp(P_0 + P_1N_t + P_2N_t^2 + P_3N_t^3)}{1 + \exp(P_0 + P_1N_t + P_2N_t^2 + P_3N_t^3)}$$

where N_t is the initial prey density, N_a is the number of preys consumed, and N_a/N_t is the probability of being eaten. P_0, P_1, P_2, and P_3 are the intercept, linear, quadratic, and cubic coefficients, respectively, estimated using the maximum likelihood method. Type-II functional response is described by $P_1 < 0$, as the proportion of prey consumed, declines gradually with increasing the initial prey density offered. If $P_1 > 0$ and $P_2 < 0$, the proportion of prey consumed is positively density-dependent, i.e., prey consumption increases with an increase in prey density, which indicates a Type-III functional response (Juliano, 2001).

The data were then transformed using the Michaelis–Menten equation as shown below (Lampert and Sommer, 1997; Sarma and Nandini, 2002)

$$V = Vmax \times S(S + Kt)$$

Where, V = prey consumption rate, V_{max} = saturation value of consumption rate, S = density of prey offered k_t = the half saturation constant, the prey concentration at which $^{V_{max}}/_{2}$ is reached. Functional response curves were plotted using Excel. A Shapiro–Wilk test was used to assess the normal distribution.

Results

The presence of MPs significantly affected the functional response curves of adult *P. annandalei* feeding on either exclusive ciliate or exclusive rotifer diets. In general, *P. annandalei* adults exhibited a classical Type-II functional response to ciliates or rotifers when offered as monospecific or in combination with algae (Figure 1). Where no clear asymptote was observed, the proportion of prey consumed in relation to initial prey density declined significantly in the MP-contaminated medium (Figure 2). The prey ingestion rate initially increased with increasing prey density but reached a plateau, after which the prey ingestion rates slightly declined (Figures 1A, B). The maximum number of rotifers consumed at the highest prey density (160 ind. 100 mL^{-1}) in medium without MP was ~89.2 ± 5.51, while the maximum number of rotifers consumed by three adults *P. annandalei* at the highest prey density was (~40.4 ± 4.26) in the MP contaminated medium. The same prey species and density when offered in combination with MP showed a deviation from the Type-II functional response for copepodites (Figure 2), as the gradual decrease in the proportion of prey consumed was not significant (Pearson's correlation; $R^2 = 0.018$; $P = 0.086$) with increasing prey concentration (Figure 3) for copepodites (C3–C4).

The copepodites (C3–C4) of *P. annandalei* exhibited different responses to the prey type and density. The functional response curve for copepodites with ciliate diet showed a Type-II curve, where the rate of prey consumption increased with an increase in prey density but eventually levelled off and declined thereafter. Type-II functional

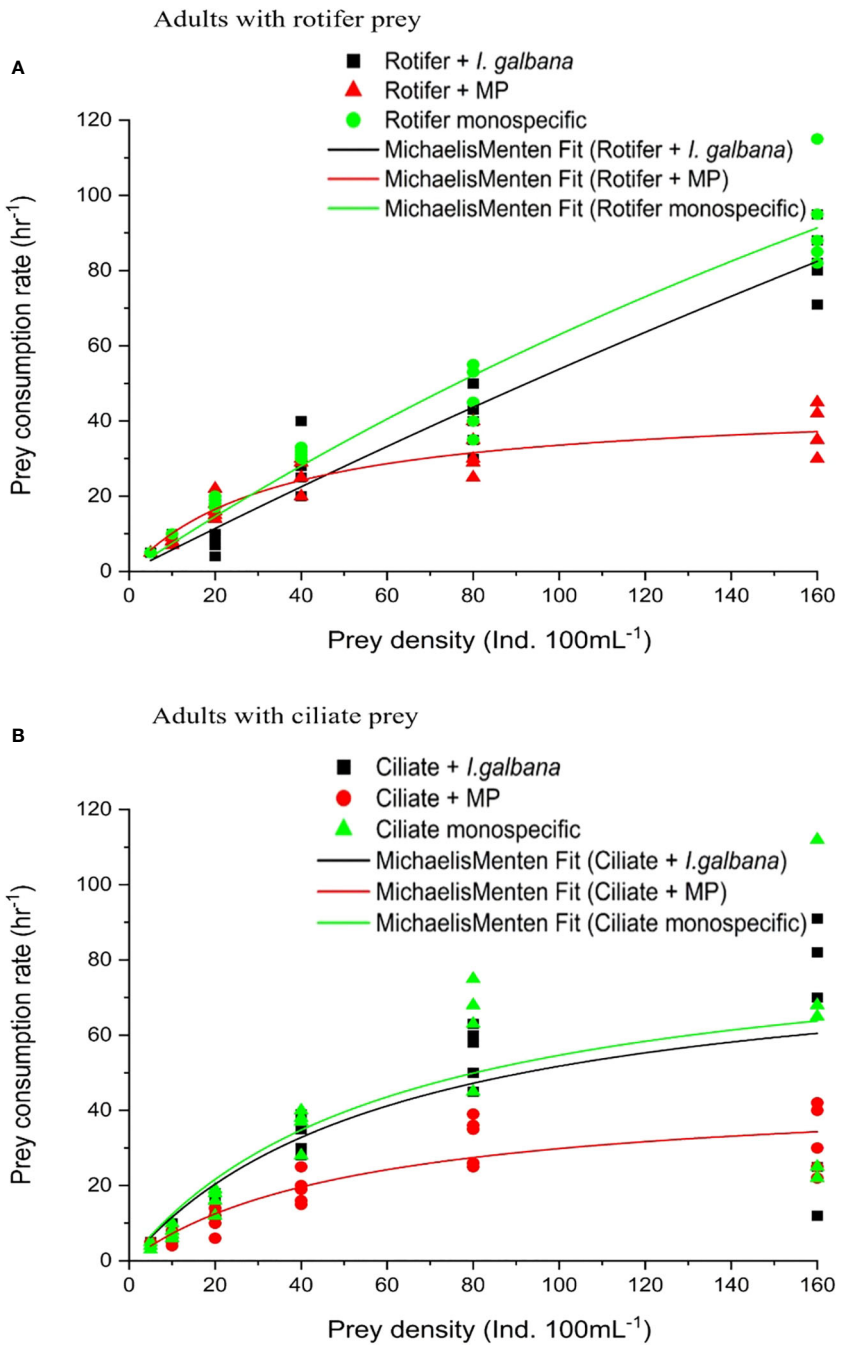


FIGURE 1
Functional response of *P. annandalei* adults offered monospecific prey, in the presence of microalgae *I. galbana*, and in the presence of Microplastic particles; **(A)** Rotifer prey *B. rotundiformis*, **(B)** Ciliate prey *Frontonia microstoma*. Shown are the actual prey consumption and the transformation of the curve using the Michaelis–Menten equation.

responses of copepodites were recorded with monospecific rotifer and ciliate prey types (Figure 3). In the presence of microalgae, copepodites showed a stronger response to ciliate density (R^2 = 0.77; P <0.001), exhibiting a Type-II functional response (Figure 3). The presence of MPs along with natural prey affected prey consumption in both adults and copepodites. The functional response curve of copepodites showed a Type-II functional response with rotifers both with a monospecific diet and with algae (Figure 2A), but the functional response curve in the presence of MPs showed linear regression with a negative slope (Figure 3A). The gradual decline in the proportion of prey consumed showed a highly significant correlation with the increasing initial prey density (Figure 4). Copepodites did not exhibit any response with rotifers or ciliates to the changed initial prey availability in the medium treated with MP (Figure 3). In the MP medium, the proportion of initial prey consumed by the copepodites did not correlate significantly (R^2 = 0.02, MP-rotifer; R^2 = 0.2, MP-ciliate) with increasing initial prey density (Figure 4).

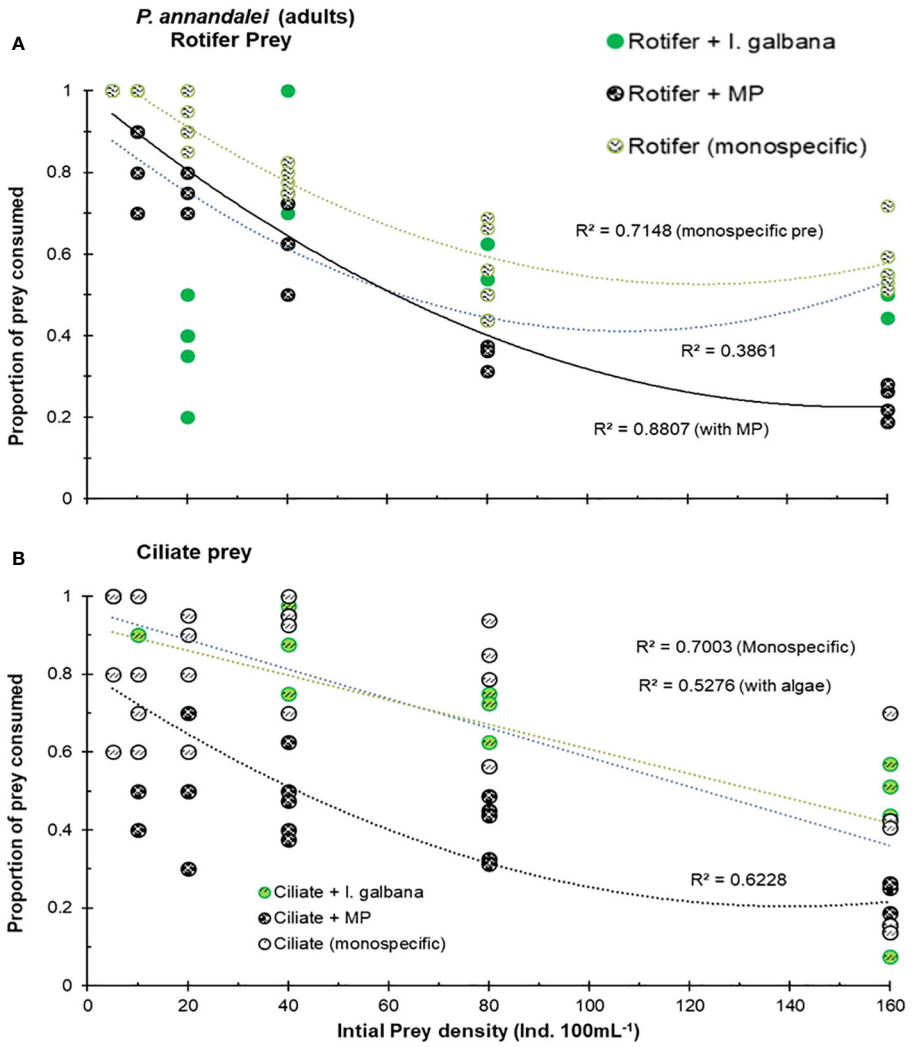


FIGURE 2
Proportion of initial prey consumed by the copepod *P. annandalei* adults, in the medium with monospecific prey, in the presence of microalgae *I. galbana*, and in the presence of MP particles **(A)** Rotifer prey *B. rotundiformis*, **(B)** Ciliate prey *Frontonia microstoma*.

In the present study, adults in all cases ingested natural rotifer and ciliate prey, but the presence of MPs differentially affected ingestion depending on food type. Regardless of prey density, the prey ingestion rates were significantly lower (Figure 5; $p < 0.01$; one-way ANOVA) for rotifer prey than for ciliate prey. The ingestion of natural prey by *P. annandalei* adults and copepodites varied significantly with prey density, MPs concentration, and their interaction (Figure 5; $p < 0.001$; two-factor ANOVA). MP contamination did not affect prey ingestion in adults or copepodites at lower prey densities. The number of preys consumed in case of rotifer diet was maximum at the maximum prey density (160 ind. 100 mL^{-1}) ~80.4 ± 2.7 which reduced by ~60% in presence of MPs. A similar decrease in prey consumption at maximum density was observed in the case of ciliate *F. microstoma* (maximum prey consumed with only ciliate prey was ~117 ± 4.55, while in the presence of MPs, ciliate consumption was reduced to ~39.8 ± 2.305 (Figure 5). At medium and high prey densities, we recorded a 40%–66% reduction in ciliate ingestion and 20%–60% reduction in rotifer ingestion by *P. annandalei* adults (Figure 5; Table 2). In the case of copepodites, we recorded a 40% to 98% reduction in ciliate consumption and a 50% to 91% reduction in rotifer consumption due to the presence of MP in the medium (Figure 5; Table 2).

Discussion

The main aim of this study was to obtain definitive information on the impact of MP contamination on feeding rates in *P. annandalei*, a common brackish water copepod species, with natural prey. In the natural setup, the distribution of prey species and MP (Rajan et al., 2023) are patchy, where copepods encounter an array of prey concentrations (Hwang et al., 2010). Various types of MP, similar to the size range of microzooplankton, are abundant in coastal and estuarine ecosystems (Cole et al., 2019; Rajan et al., 2023).

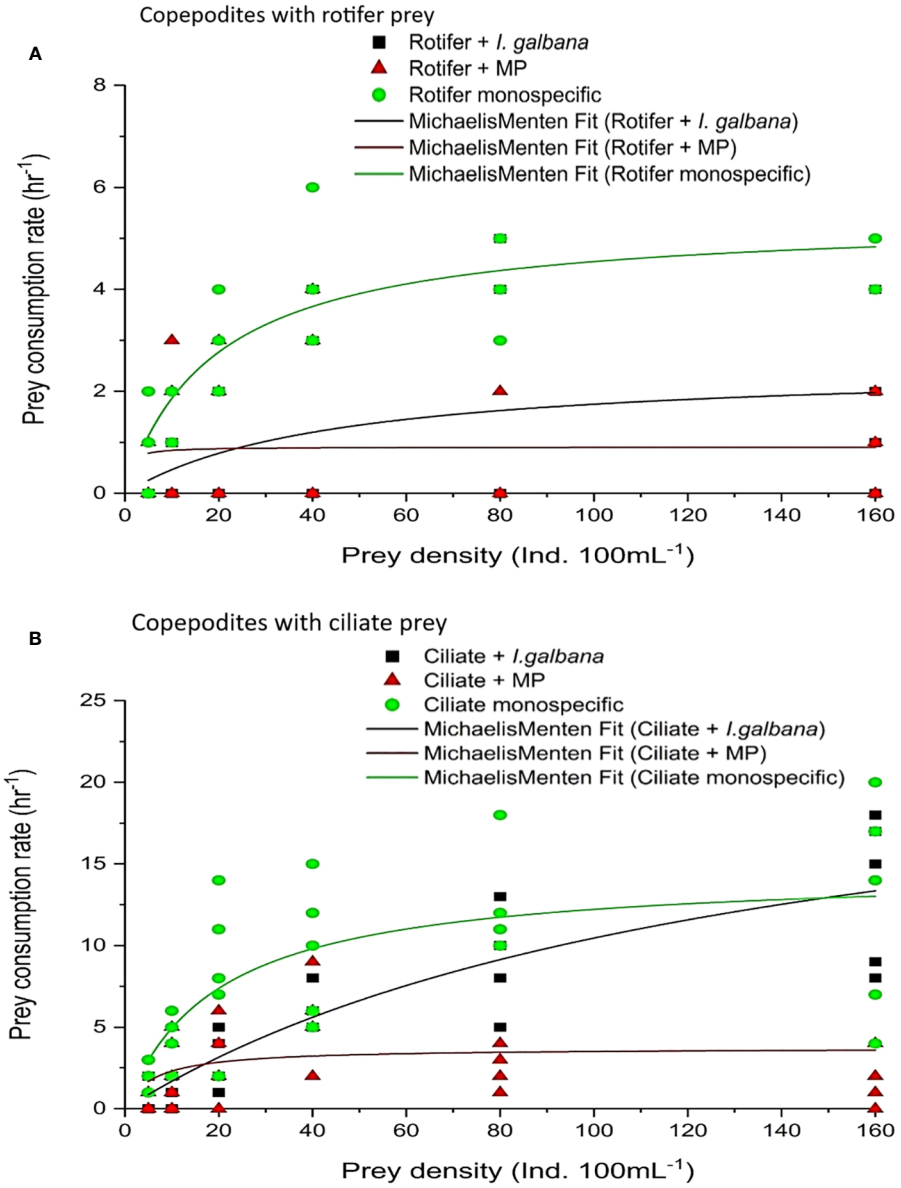


FIGURE 3
Functional response of *P. annandalei* copepodites (C3–C4) offered monospecific prey, in the presence of microalgae *I. galbana*, and in the presence of Microplastic particles; **(A)** Rotifer prey *B. rotundiformis*, **(B)** Ciliate prey *Frontonia microstoma*. Shown are the actual prey consumption and the transformation of the curve using the Michaelis–Menten equation.

Evaluation of MP imposed alterations in consumer behavior is complex because of (i) patchy distribution and natural concentrations of MP, (ii) limited information on MP concentration-specific ingestion of natural diet, and (iii) lack of quantitative data on behavioral responses of zooplankton at varying MP concentrations. Unless colonized by biota or affected by external pollutants, MP are inert particles, and the main mode of effects on consumers is food dilution and behavioral responses, such as functional and numerical responses of consumers to MP, as the polymer structure and chemical properties of particles are less important (Mehinto et al., 2022). While food dilution outweighs other effects, the major response of organisms with a dietary niche size overlapping with MP would be the adjustment of feeding rates, i.e., functional response.

Our results on prey ingestion by *P. annandalei* adults and copepodites were within the ranges reported for related taxa (Hwang et al., 2009; Dhanker et al., 2012; He et al., 2020). Recent studies have also confirmed ingestion of MP by copepods (Desforges et al., 2015; Cheng et al., 2020; Bai et al., 2021); however, this is the first study to understand the effect of MP on natural prey ingestion in *P. annandalei*. Contributing to the size range, sorption potential, hydrophobicity, wide occurrence in the global ocean (Foekema et al., 2013; Mathalon and Hill, 2014; Ugwu et al., 2021), and bioavailability to zooplankton (Cole et al., 2013; Setälä et al., 2014), ichthyoplankton (Rochman et al., 2013; Batel et al., 2016), and other meso-macro-sized invertebrates (Browne et al., 2008; Graham and Thompson, 2009; Gray and Weinstein, 2017; Kumari et al., 2023a), MPs are likely to have large-scale effects

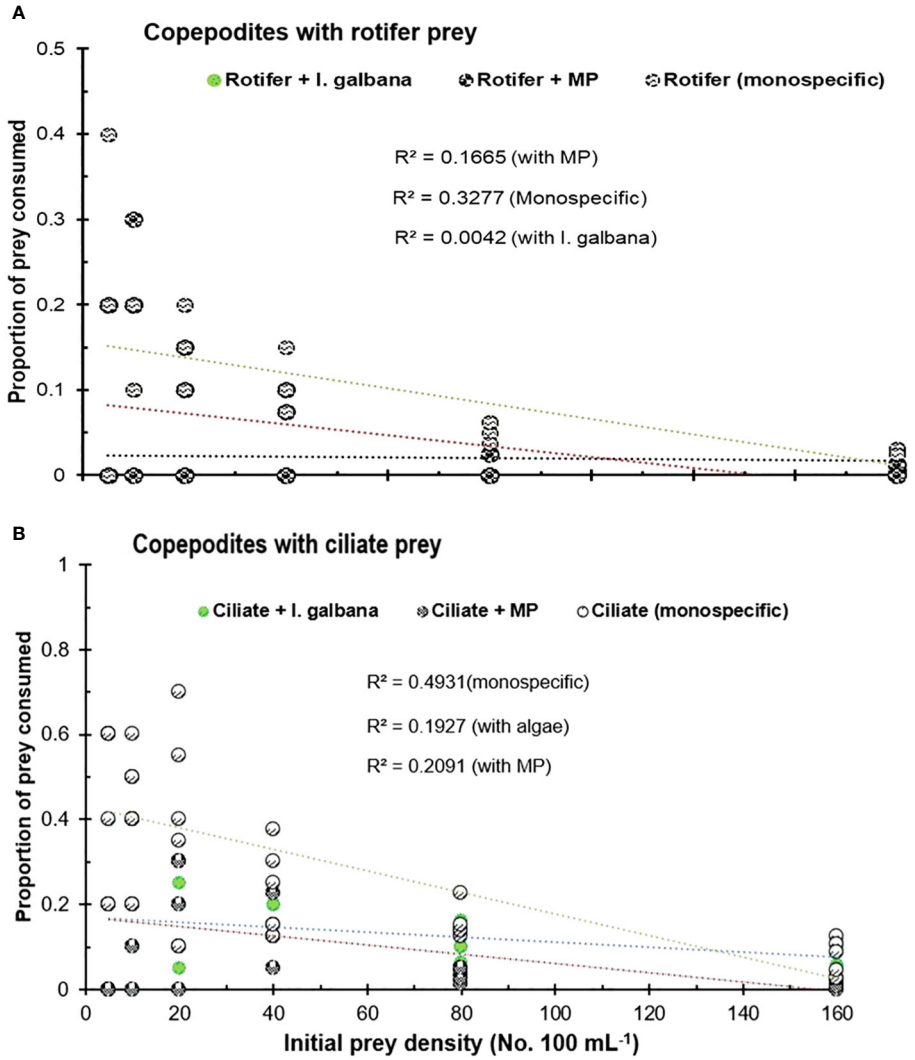


FIGURE 4
Proportion of initial prey consumed by the copepodid (C3–C4) stage of P. *annandalei* in the medium with monospecific prey, in the presence of microalgae *I. galbana*, and in the presence of MP particles. **(A)** Rotifer prey *B. rotundiformis*. **(B)** Ciliate prey *Frontonia microstoma*.

on marine ecosystem functioning. Various experimental results have revealed lethal and sublethal effects on different organisms (von Moos et al., 2012; Rochman et al., 2013; Au et al., 2015; Gray and Weinstein, 2017).

P. annandalei showed a direct functional response to varying concentration of prey in the medium. Thus, the presence of MPs (size range 15 µm–30 µm diameter) in the environment of *P. annandalei* directly affects the functional response and feeding behavior of both adults and the developmental stages of copepods. The prey ingestion rate showed a decline in the presence of MPs in all cases, suggesting interference of MPs in the feeding behavior and feeding efficiency of *P. annandalei*. Copepods can ingest plastics by either selective grazing or by indiscriminately ingesting particles with natural prey (Traboni et al., 2023). Copepods can also actively refuse MPs or passively reduce MP intake by physical impacts/interference or chemical toxicity associated with plastics (leachates or sorbed pollutants) (Cássio et al., 2022). Similarly, the presence of plastics in the environment, along with natural prey, may obstruct predator–prey interactions, leading to a reduced prey ingestion rate (Turner and Tester, 1989). The linear increase in prey ingestion by copepods with increasing prey density suggests a Type-I functional response, which assumes a linear increase in food ingestion rate with food concentration, either for all initial food concentrations or only for the particular maximum concentration, beyond which the feeding rate is constant. This could be attributed to the adaptive feeding response of *P. annandalei*, as has also been reported for other copepods (Abrams, 1990; Kumar and Rao, 1999; Kerfoot and DeMott, 2019). The MP incurred reduction in ciliate and rotifer prey ingestion, recorded in this study, indicating the adaptive feeding strategy, as adults of *P. annandalei* did not exhibit any discernible reduction in prey consumption at lower concentrations; however, at higher concentrations, it showed a 20% to 60% reduction in rotifer ingestion and a 40% to 66% reduction in ciliate ingestion. Thus, the presence of a more satiating prey has the potential to reduce MP uptake by copepods. *P. annandalie* is

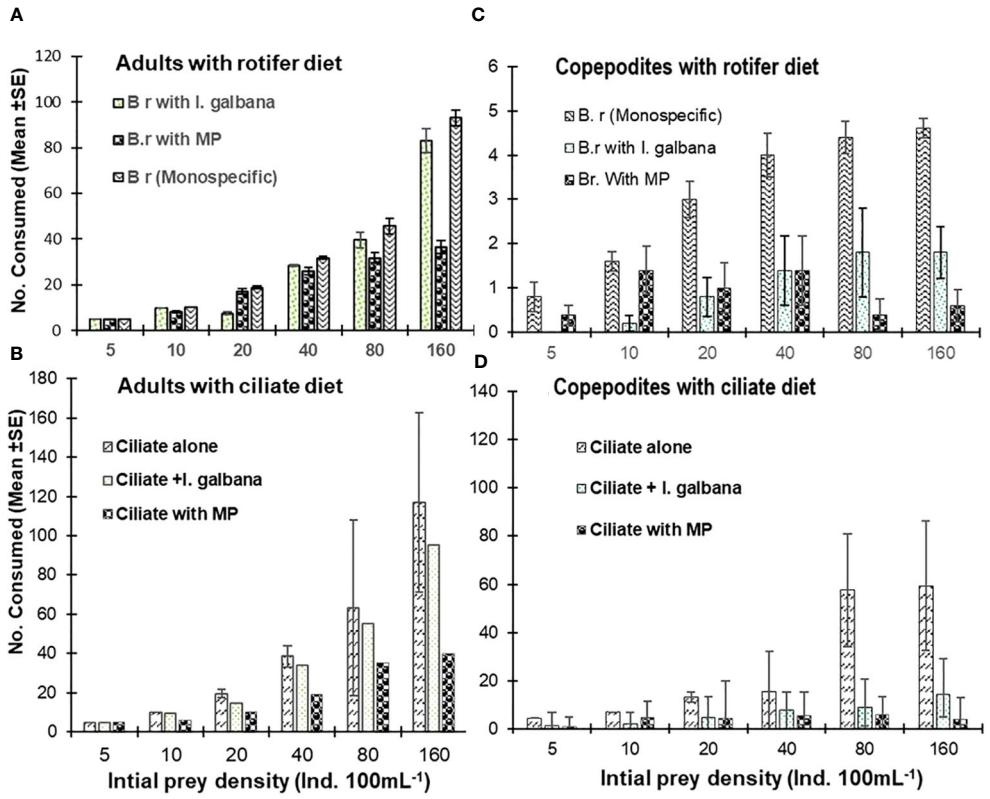


FIGURE 5
No. of prey consumed (mean ± SD) by the adults **(A, B)**, and copepodites **(C, D)** of *Pseudodiaptomus annandalei* when offered as prey the rotifer *Brachionus rotundiformis* **(A, C)** and the ciliate *Frontonia microstoma* **(B, D)** as monospecific prey, in the medium containing either microalgae *I. galbana* or MP particle.

mainly a suspension feeder copepod, though occasionally capture a food item through a surprise jerky movement of appendages. A previous study recorded different feeding modes for *P. annandalei* when dealing with different prey sizes and mobilities (Dhanker et al., 2013). MP-driven alterations in the behavioral and physiological responses of the calanoid copepods *Acartia clausi* and *C. typicus* have been attributed to differences in the feeding strategies of the two species (Svetlichny et al., 2021). Further comparative studies are required to understand the role of different feeding strategies in the uptake of MP by *P. annandalei*.

In the present study, copepods showed neither linear nor curvilinear relationships with the initial prey concentrations in the medium with MP. Conclusively *P. annandalei* was much less efficient in feeding *B. rotundiformis* or *F. microstoma* when MP was present, indicating a negative impact of MPs contamination. The ability of copepods, such as *Temora longicornis*, to selectively refuse MPs and their preference for natural prey has been previously demonstrated (Xu et al., 2022). Overall, copepods have been shown to be equipped with diverse sensors on their antennae, feeding appendages or body surfaces for detecting either hydromechanical or chemical signals emitted by their prey (Turner and Tester, 1989; Vanderploeg et al., 1990; DeMott, 1995). The reduced frequency of prey predator encounters in the presence of MPs may be a factor in changes in the functional response. In addition, accidental or

TABLE 2 MP induced changes in prey ingestion rates on natural prey (*F. microstoma* and *B. rotundiformis*) by adult and copepodid stages of *P. annandalei* at different food density levels.

Alternative in the medium	Copepod stages used	% change in ciliate ingestion			% change in rotifer ingestion		
		Food density			Food density		
		Low	Medium	High	Low	Medium	High
Microalgae	Adults	_−2.00	−18.20	−15.70	−2.00	−5.34	−11.85
	Copepodites	−69.81	−56.93	−80.41	−93.75	−69.17	−60.00
Microplastic	Adults	−21.00	−48.56	−55.30	−10.00	−14.22	−45.56
	Copepodites	−40.80	−66.03	−91.77	−31.25	−65.83	−88.93

voluntary intake of MPs by *P. annandalei* could induce false feelings of satiety that may reduce the hunger level of copepods (Gommer et al., 2018; Botterell et al., 2019). MPs acceptance and rejection by copepods create a challenging environment for copepods to extract natural food particles from the environment because of the morphological similarity of MPs with the natural diet (Nielsen et al., 2020) and microbial coating on MP surfaces as biofilms (Zimmermann et al., 2019). The purpose of this study was to evaluate the effect of MP on the functional response (prey consumption rate) of *P. annandalei*. The complete life cycle of this species has been studied on an exclusive algal diet (Beyrend-Dur et al., 2011; Nielsen et al., 2020), and lower availability of microalgae showed an increase in MP ingestion and retention in *P. annandalei* (Cheng et al., 2020); therefore, it is also relevant to study the effect of MP on the clearance rate.

Earlier reports suggest ingestion of inorganic particles of low nutritional value (Arendt et al., 2011) resulting in pseudo-satiation (Turner and Tester, 1989). However, a lack of nutrients adversely affects an organism's reproductive potential and inclusive fitness (White and Dagg, 1989; Kang, 2012).

The brackish water, egg-carrying euryhaline copepod, *P. annandalei*, is distributed widely in tropical and subtropical Indo-Pacific with an ability to survive in altered marine and brackish waters (Beyrend-Dur et al., 2011; Beyrend-Dur et al., 2013; Blanda et al., 2015). This species is widely used as a live feed in the aquaculture industry; however, the harvest of *P. annandalei* in aquaculture ponds has declined in recent years (Blanda et al., 2015). The present results emphasize the need to assess MP contamination in aquaculture ponds and further large-scale studies of the numerical response of *P. annandalei* in the presence and absence of MP. The impact of MPs on the functional responses of *P. annandalei* is of particular concern because they have the potential for trophic transfer to fish and humans (Lee et al., 2010; Nelms et al., 2018; Ugwu et al., 2021). Furthermore, *P. annandalei* which is a filter-feeder, is more vulnerable to MPs ingestion. It is important to consider that a cocktail of various additives and chemical contaminants (Rochman et al., 2019; Zimmermann et al., 2019) are adsorbed by plastic surfaces and ingested along with microplastics, which further adds to toxicity and oxidative stress (Jeong et al., 2017; Qiao et al., 2019). Although the present study does not suggest any toxic effects of MPs, the effects on functional responses are suggestive of chronic effects on ecosystem functioning of planktonic organisms, particularly at the base of the pelagic food webs, hints at the knock-on alterations in ecosystem functioning, resulting in trophic cascades (Galloway et al., 2017; Botterell et al., 2019; Nair and Perumal, 2022). MPs are expected to become more prevalent in aquatic environments because of their extensive use and dumping in aquatic ecosystems (Zeynep et al., 2023). Their potential to transport and adsorb toxic materials, as well as their absorbance by living cells, tissues, and organs of aquatic organisms, and interfere with physiological functions, all cause significant health risks. Could the observed MP imposed alterations in functional responses have implications for the numerical responses of copepods? The predator's functional response is hypothesized to be an important determinant of the population structure and numerical responses (Dhanker et al., 2012; Dhanker et al., 2015).

To support the generalizability of the present results, we recommend studying a range of brackish water zooplankton species used as live feed in the aquaculture industry, by applying different types of MP particles.

Conclusions

This paper reports significant changes in the functional response curve of the adult and copepodid stages of *P. annandalei* when MPs are present along with natural prey. These results clearly suggest that MP contamination adversely affects the feeding rates of *P. annandalei* adults and juveniles fed on their natural diets. The impact of MPs on the functional responses of *P. annandalei* and prey ingestion rates is of particular concern because they have the potential for trophic transfer to fish and humans. Ingestion of inert particles, such as MP, gives a sense of false satiety, which in turn results in an altered prey selection pattern (Gommer et al., 2018) and reduced saturation level prey intake. Further research is needed to fully understand the feeding modes in zooplankton species, prey behavior, prey selection patterns, and trophic transfer to fish in MP-contaminated environments.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Ethics statement

The manuscript presents research on animals that do not require ethical approval for their study.

Author contributions

JS: Conceptualization, Writing - review & editing. MR: Writing – original draft, Writing – review & editing. RG-o: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors thank the guest editor for inviting us to submit this special issue. We thank Silliman University for providing the facility for conducting this research. MR thanks Delhi Technological University for the internship.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2024.1277332/full#supplementary-material

References

Abrams, P. A. (1990). The effects of adaptive behavior on the type-2 functional response. *Ecology* 71, 877–8 85. doi: 10.2307/1937359

Akdogan, Z., Guven, B., and Kideys, A. E. (2023). Microplastic distribution in the surface water and sediment of the Ergene River. *Environ. Res.* 234, 116500. doi: 10.1016/j.envres.2023.116500. ISSN 0013-9351.

Alava, J. J., Báez, M. M., McMullen, K., Tekman, M. B., Barrows, A. P. W., Bergmann, M., et al. (2023). *Ecological Impacts of Marine Plastic Pollution, Microplastics' Food web Bioaccumulation Modelling and Global Ocean Footprint: Insights into the Problems, the Management Implications and Coastal Communities Inequities. The Nippon Foundation-Ocean Litter Project, (2019-2023)* (Main Mall, Vancouver, BC, V6T 1Z4, Canada: Institute for the Oceans and Fisheries, University of British Columbia). *Working Paper* 2023 – 01.

Amélineau, F., Bonnet, D., Heitz, O., Mortreux, V., Harding, A. M. A., Karnovsky., N., et al. (2016). MP pollution in the Greenland Sea: Background levels and selective contamination of planktivorous diving seabirds. *Environ. pollut.* 219, 1131–1139. doi: 10.1016/j.envpol.2016.09.017

Arendt, K. E., Dutz, J., Jónasdóttir, S. H., Jung-Madsen, S., Mortensen, J., Møller., E. F., et al. (2011). Effects of suspended sediments on copepods feeding in a glacial influenced sub-Arctic fjord. *J. Plankton Res.* 33, 10, 1526–1537. doi: 10.1093/plankt/fbr054

Au, S. Y., Bruce, T. F., Bridges, W. C., and Klaine, S. J. (2015). Responses of *Hyalella azteca* to acute and chronic microplastic exposures. *Environ. Toxicol. Chem.* 34, 2564–2572. doi: 10.1002/etc.3093

Auta, H. S., Emenike, C. U., and Fauziah, S. H. (2017). Distribution and importance of MPs in the marine environment: A Review of the Sources, Fate, Effects, and Potential Solutions. *Environ. Int.* 2017, 165–176. doi: 10.1016/j.envint.2017.02.013

Avignon, G. D., Hsu, S. S. H., Gregory-Eaves, I., and Ricciardi, A. (2023). Feeding behavior and species interactions increase the bioavailability of microplastics to benthic food webs. *Sci. Total Environ.* 896, 165261. doi: 10.1016/j.scitotenv.2023.165261

Ayukai, T. (1987). Discriminate feeding of the calanoid copepod *Acartia clausi* in mixtures of phytoplankton and inert particles. *Mar. Biol.* 94, 579–587. doi: 10.1007/BF00431404

Baalkhuyur, F. M., Dohaish, E.-J. A., Elhalwagy, M. E. A., Alikunhi, N. M., Suwailem, A. M., Røstad, A., et al. (2018). MP in the gastrointestinal tract of fishes along the Saudi Arabian Red Sea coast. *Mar. pollut. Bull.* 131, 407–415. doi: 10.1016/j.marpolbul.2018.04.040

Bai, Z., Wang, N., and Wang, M. (2021). Effects of microplastics on marine copepods. *Ecotoxicol Environ. Saf.* 217, 112243. doi: 10.1016/j.ecoenv.2021.112243

Barnes, D. K. A., Galgani, F., Thompson, R. C., and Barlaz, M. (2009). Accumulation and fragmentation of plastic debris in global environments. *Phil. Trans. R. Soc B* 364, 1526. doi: 10.1098/rstb.2008.0205

Batel, A., Linti, F., Scherer, M., Erdinger, L., and Braunbeck, T. (2016). Transfer of benzo[a]pyrene from microplastics to Artemia nauplii and further to zebrafish via a trophic food web experiment: CYP1A induction and visual tracking of persistent organic pollutants. *Environ. Toxicol. Chem.* 35 (7), 1656–1666. doi: 10.1002/etc.3361

Beaugrand, G., Brander, K. M., Alistair, J. A., Souissi, S., and Reid, P. C. (2003). Plankton effect on cod recruitment in the North Sea. *Nature* 426, 661–664. doi: 10.1038/nature02164

Beyrend-Dur, D., Kumar, R., Rao, T. R., Souissi, S., Cheng, S. H., and Hwang, J. S. (2011). Demographic parameters of adults of *Pseudodiaptomus annandalei* (Copepoda: Calanoida): Temperature–salinity and generation effects. *J. Exp. Mar. Biol. Ecol.* 404 (1–2), 1–14. doi: 10.1016/j.jembe.2011.04.012

Beyrend-Dur, D., Souissi, S., and Hwang, J. S. (2013). Population dynamics of calanoid copepods in the subtropical mesohaline Danshuei Estuary (Taiwan) and typhoon effects. *Ecol. Res.* 28, 771–780. doi: 10.1007/s11284-013-1052-y

Bhuyan, M. (2022). Effects of MPs on fish and in human health. *Front. Environ. Sci.* 10, 827289. doi: 10.3389/fenvs.2022.827289

Blanda, E., Drillet, G., Huang, C.-C., Hwang, J.-S., Jakobsen, H. H., Rayner, T. A., et al. (2015). Trophic interactions and productivity of copepods as live feed from tropical Taiwanese outdoor aquaculture ponds. *Aquaculture* 445, 11–21. doi: 10.1016/j.aquaculture.2015.04.003

Botterell, Z. L. R., Beaumont, N., Dorrington, T., Steinke, M., Thompson, R. C., and Lindeque, P. K. (2019). Bioavailability and effects of microplastics on marine zooplankton: A review. *Environ. pollut.* 245, 98–110. doi: 10.1016/j.envpol.2018.10.065

Browne, M. A., Dissanayake, A., Galloway, T. S., Lowe, D. M., and Thompson, R. C. (2008). Ingested microscopic plastic translocates to the circulatory system of the mussel, Mytilus edulis (L). *Environ. Sci. Technol.* 42 (13), 5026–5031. doi: 10.1021/es800249a

Bundy, M. H., Gross, T. F., Vanderploeg, H. A., and Strickler, J. R. (1998). Perception of inert particles by calanoid copepods: Behavioral observations and a numerical model. *J. Plankton Res.* 20, 2129–2152. doi: 10.1093/plankt/20.11.2129

Bundy, M. H., and Vanderploeg, H. A. (2002). Detection and capture of inert particles by calanoid copepods: the role of the feeding current. *J. Plankton Res.* 24, 3, 215–223. doi: 10.1093/plankt/24.3.215

Cássio, F., Batista, D., and Pradhan, A. (2022). Plastic interactions with pollutants and consequences to aquatic ecosystems: what we know and what we do not know. *Biomolecules* 12 (6), 798. doi: 10.3390/biom1206079898

Chan, K., Boutin, S., Hossie, T. J., Krebs, C. J., O'Donoghue, M., and Murray, D. L. (2017). Improving the assessment of predator functional responses by considering alternate prey and predator interactions. *Ecology* 98, 1787–17 96. doi: 10.1002/ecy.1828

Cheng, Y., Wang, J., Yi, X., Li, L., Liu, X., and Ru, S. (2020). Low microalgae availability increases the ingestion rates and potential effects of microplastics on marine copepod *Pseudodiaptomus annandalei*. *Mar. Poll. Bul.* 152, 110919. doi: 10.1016/j.marpolbul.2020.110919

Cole, M., Coppock, R., Lindeque, P. K., Altin, D., Reed, S., Pond, D. W., et al. (2019). Effects of nylon MP on feeding, lipid accumulation, and moulting in a coldwater copepod. *Environ. Sci. Technol.* 53, 7075–7082. doi: 10.1021/acs.est.9b01853

Cole, M., Lindeque, P., Fileman, E., Halsband, C., and Galloway, T. S. (2015). The impact of polystyrene MPs on feeding, function and fecundity in the marine copepod *Calanus helgolandicus*. *Environ. Sci. Technol.* 49 (2), 1130–1137. doi: 10.1021/es504525u

Cole, M., Lindeque, P., Fileman, E., Halsband, C., Goodhead, R., Moger, J., et al. (2013). Microplastic ingestion by zooplankton. *Environ. Sci. Technol.* 47 (12), 6646–6655. doi: 10.1021/es400663f

Cowles, T. J., Desiderio, R. A., and Carr, M.-E. (1998). Small-scale planktonic structure: persistence and trophic consequences. *Oceanography* 11 (1), 4–9. doi: 10.5670/oceanog.1998.08

Cunningham, E. M., Cuthbert, R. N., Coughlan, N. E., Kregting, L., Cairnduff, V., and Dick, J. T. A. (2021). Microplastics do not affect the feeding rates of a marine predator'. *Sci. Total Environ.* 779, 146487. doi: 10.1016/j.scitotenv.2021.146487

Cunningham, E. M., Kiriakoulakis, K., Dick, J. T. A., Kregting, L., Schuchert, P., and Sigwart, J. D. (2020). Driven by speculation, not by impact - the effects of plastic on fish species. *J. Fish Biol.* 96 (6), 1294–1297. doi: 10.1111/jfb.14303

DeMott, W. R. (1986). The role of taste in food selection by freshwater zooplankton. *Oecologia* 69, 334–340. doi: 10.1007/BF00377053

DeMott, W. R. (1988). Discrimination between algae and artificial particles by freshwater and marine copepods. *Limnol. Oceanogr.* 33, 397–408. doi: 10.4319/lo.1988.33.3.0397

DeMott, W. R. (1995). Optimal foraging by a suspension-feeding copepod: responses to short-term and seasonal variation in food resources. *Oecologia* 103 (2), 230–240. doi: 10.1007/BF00329085

DeMott, W. R., and Moxter, F. (1991). Foraging cyanobacteria by copepods: responses to chemical defense and resource abundance. *Ecology* 72 (5), 1820–1834. doi: 10.2307/1940981

Desforges, J. P., Galbraith, M., and Ross, P. S. (2015). Ingestion of microplastics by zooplankton in the Northeast Pacific Ocean. *Arch. Environ. Contam Toxicol.* 69 (3), 320–330. doi: 10.1007/s00244-015-0172-5

Dhanker, R., Kumar, R., and Hwang, J. S. (2012). Predation by *Pseudodiaptomus annandalei* (Copepoda: Calanoida) on rotifer prey: size selection, egg predation and effect of algal diet. *J. Exp. Mar. Biol. Ecol.* 414–415, 44–53. doi: 10.1016/j.jembe.2012.01.011

Dhanker, R., Kumar, R., Tseng, L. C., and Hwang, J. S. (2013). Ciliate (*Euplotes* sp.) predation by *Pseudodiaptomus annandalei* (Copepoda: Calanoida) and the effects of mono-algal and pluri-algal diets. *Zool. Stud.* 52, 34. doi: 10.1186/1810-522X-52-34

Dhanker, R., Molinero, J. C., Kumar, R., Tseng, L. C., Ianora, A., and Hwang, J. S. (2015). Responses of the estuarine copepod *Pseudodiaptomus annandalei* to diatom polyunsaturated aldehydes: Reproduction, survival and postembryonic development. *Harmful Algae* 43, 74–81. doi: 10.1016/j.hal.2015.02.002

Enríquez-García, C., Nandini, S., and Sarma, S. S. S. (2013). Feeding behaviour of Acanthocyclops americanus (Marsh)(Copepoda: Cyclopoida). *J. Nat. Hist.* 47 (5-12), 853–862. doi: 10.1080/00222933.2012.747637

Ferrao Filho, A. D. S., Demott, W. R., and Tessier, A. J. (2005). Responses of tropical cladocerans to a gradient of resource quality. *Freshwater Biol.* 50 (6), 954–964.

Foekema, E. M., De Gruijter, C., Mergia, M. T., van Franeker, J. A., Murk, A. J., and Koelmans, A. A. (2013). Plastic in North Sea fish. *Environ. Sci. Technol.* 47 (15), 8818–8824. doi: 10.1021/es400931b

Franzellitti, S., Canesi, L., Auguste, M., Wathsala, R. H., and Fabbri, E. (2019). Microplastic exposure and effects in aquatic organisms: A physiological perspective. *Environ. Toxicol. Pharmacol.* 68, 37–51. doi: 10.1016/j.etap.2019.03.009

Frost, W. E. (1977). The food of Charr *Salvelinus willughbii* (Gunther) in Windermere. *J. Fish Biol.* 11, 531–547. doi: 10.1111/j.1095-8649.1977.tb05710.x

Fulton, R. S., and Paerl, H. W. (1987). Toxic and inhibitory effects of the blue-green alga *Microcystis aeruginosa* in herbivorous zooplankton. *J. Plankton Res.* 9, 837–855. doi: 10.1093/plankt/9.5.837

Galloway, T., Cole, M., and Lewis, C. (2017). Interactions of microplastic debris throughout the marine ecosystem. *Nat. Ecol. Evol.* 1, 116. doi: 10.1038/s41559-017-0116

Gommer, R., Bom, R. A., Fijen, T. P. M., and van Gils, J. A. (2018). Stomach fullness shapes prey choice decisions in crab plovers (*Dromas ardeola*). *PloS One* 13 (4), e0194824. doi: 10.1371/journal.pone.0194824

Graham, E. R., and Thompson, J. T. (2009). Deposit and suspension- feeding sea cucumbers (Echinodermata) ingest plastic fragments. *J. Exp. Mar. Biol. Ecol.* 368, 22–29. doi: 10.1016/j.jembe.2008.09.007

Gray, A. D., and Weinstein, J. E. (2017). Size- and shape-dependent effects of microplastic particles on adult daggerblade grass shrimp (Palaemonetes pugio). *Environ. Toxicol. Chem.* 36 (11), 3074–3080. doi: 10.1002/etc.3881

Gunaalan, K., Almeda, R., Lorenz, C., Vianello, A., Iordachescu, L., Papacharalampos, K., et al. (2023). Abundance and distribution of microplastics in surface waters of the Kattegat/ Skagerrak (Denmark). *Environ. pollut.* 318, 120853. doi: 10.1016/j.envpol.2022.120853

Halstead, J. E., Smith, J. A., Carter, E. A., Lay, P. A., and Johnston, E. L. (2018). Assessment tools for MPs and natural fibres ingested by fish in an urbanised estuary. *Environ. pollut.* 234, 552–561. doi: 10.1016/j.envpol.2017.11.085

Hartmann, H. J., Taleb, H., Aleya, L., and Lair, N. (1993). Predation on ciliates by the suspension-feeding calanoid copepod *Acanthodiaptomus denticornis*. *Can. J. Fish. Aquat. Sci.* 50, 1382–1393. doi: 10.1139/f93-158

He, X., Cheng, F., and Chen, W. (2020). Effects of maternal phosphorus deficiency on reproductive allocation of the copepod *Pseudodiaptomus annandalei* Sewel. *Hydrobiologia* 847, 2035–2048. doi: 10.1007/s10750-020-04229-7

Holling, C. S. (1959). The components of predation as revealed by a study of small-mammal predation of the European pine sawfly. *Can. Entomologist.* 91 (5), 293–320. doi: 10.4039/Ent91293-5

Hwang, J. S., Kumar, R., Hsieh, C. W., Kuo, A. Y., Souissi, S., Hsu, M. H., et al. (2010). Patterns of zooplankton distribution along the marine, estuarine and riverine portions of the Danshuei ecosystem in northern Taiwan. *Zoological Studies* 49 (3), 335–352.

Hwang, J. S., Kumar, R., and Kuo, C.-S. (2009). Impacts of predation by the copepod, mesocyclops pehpeiensis, on life table demographics and population dynamics of four cladoceran species: A comparative laboratory study. *Zoological Stud.* 48 (6), 738–752.

Isinibilir, M., Eryalçın, K. M., and Kideys, A. E. (2023). Effect of Polystyrene Microplastics in Different Diet Combinations on Survival, Growth and Reproduction Rates of theWater Flea (*Daphnia magna*). *Microplastics* 2, 27–38. doi: 10.3390/microplastics2010002

Isinibilir, M., Svetlichny, L., Mykitchak, T., Türkeri, E. E., Eryalçın, K. M., Doğan, O., et al. (2020). Microplastic consumption and its effect on respiration rate and motility of *calanus helgolandicus* from the Marmara Sea. *Front. Mar. Science.* 7. doi: 10.3389/fmars.2020.603321

Iyer, N., and Rao, T. R. (1996). Responses of the predatory rotifer *Asplanchna intermedia* to prey species differing in vulnerability: laboratory and field studies. *Freshw. Biol.* 36 (3), 521–533. doi: 10.1046/j.1365-2427.1996.d01-510.x

Jabeen, K., Su, L., Li, J., Yang, D., Tong, C., and Mu, J. (2017). MPs and mesoplastics in fish from coastal and fresh waters of China. *Environ. pollut.* 221, 141–149. doi: 10.1016/j.envpol.2016.11.055

Jeong, C. B., Kang, H. M., Lee, M. C., Kim, D. G., Han, J., Hwang, D. S., et al. (2017). Adverse effects of microplastics and oxidative stress-induced MAPK/Nrf2 pathway-mediated defense mechanisms in the marine copepod. *Paracyclopina nana*. *Sci. Rep.* 2017, 7. doi: 10.1038/srep41323

Juliano, S. A. (2001). Nonlinear curve fitting: predation and functional response curves. *Design Anal. Ecol. Experiments* 2, 178–96. doi: 10.1093/oso/9780195131871.003.0010

Kang, H. K. (2012). Effects of suspended sediments on reproductive responses of Paracalanus sp. (Copepoda: Calanoida) in the laboratory. *J. Plankton Res.* 34, 7, 626–635. doi: 10.1093/plankt/fbs033

Kanhai, L. D. K., Gardfeldt, K., Krumpen, T., Thompson, R. C., and O'Connor, I. (2020). MPs in sea ice and seawater beneath ice floes from the Arctic Ocean. *Sci. Rep.* 10 (1), 5004. doi: 10.1038/s41598-020-61948-6

Kerfoot, C. W., and DeMott, W. R. (2019). Food web dynamics: dependent chains and vaulting. In *Trophic interactions within aquatic ecosystems* (pp. 347–382). Routledge.

Kim, S. W., Kim, D., Chae, Y., and An, Y.–J. (2018). Dietary uptake, biodistribution, and depuration of microplastics in the freshwater diving beetle *Cybister japonicus:* Effects on predacious behavior. *Environ. Poll* 242, 839e844. doi: 10.1016/j.envpol.2018.07.071

Kumar, R. (2003). Effects of *mesocyclops thermocyclopoides* (Copepoda: cyclopoida) predation on the population growth patterns of different prey species. *J. Freshwat. Ecol.* 18 (3), 383–393. doi: 10.1080/02705060.2003.9663974

Kumar, R. (2005). Feeding modes and associated mechanisms in zooplankton, (In *ecology of plankton* ed: kumar, A). *Daya Publ.*, 241–265.

Kumar, R., and Rao, T. R. (1999). Effect of algal food on animal prey consumption rates in the omnivorous copepod, *mesocyclops thermocyclopoides*. *Internat. Rev. der gesamten Hydrobiologie und Hydrographie* 84 (5), 419–426. doi: 10.1002/iroh.199900035

Kumar, R., and Rao, T. R. (2001). Effect of the cyclopoid copepod *Mesocyclops thermocyclopoides* on the interactions between the predatory rotifer *Asplanchna intermedia* and its prey *Brachionus calyciflorus* and *B. angularis*. *Hydrobiologia* 453, 261–268. doi: 10.1023/A:1013105125857

Kumari, N., Samantaray, B. P., Patel, A., and Kumar, R. (2023a). Microplastics affect rates of locomotion and reproduction via dietary uptake in globally invasive snail. *Physa acuta. Water* 15 (5), 928. doi: 10.3390/w15050928

Kumari, N., Yadav, D. K., Khan, P. K., and Kumar, R. (2023b). Occurrence of plastics and their characterization in wild-caught fish species (Labeo rohita, Wallago attu, and Mystus tengara) of River Ganga (India) compared to a commercially cultured species (L. rohita). *Environ. pollut.* 334, 122141. doi: 10.1016/j.envpol.2023.122141

Lampert, W., and Sommer, U. (1997). *Limnoloecology: the ecology of lakes and streams* (Oxford: Oxford University Press).

Lee, C.-H., Dahms, H.-U., Cheng, S.-H., Souissi, S., Schmitt, F. G., Kumar, R., et al. (2010). Predation of *Pseudodiaptomus annandalei* (Copepoda: Calanoida) by the grouper fish fry *Epinephelus coioide*s under different hydrodynamic conditions. *J. Exp. Mar. Bio. Ecol.* 393 (1-2), 17–22. doi: 10.1016/j

Li, D., Peng, G., and Lixin, Z. (2019). Progress and prospects of marine microplastic research in China. *Anthropocene Coasts.* 2 (1), 330–339. doi: 10.1139/anc-2018-0014

Mathalon, A., and Hill, P. (2014). Microplastic fibers in the intertidal ecosystem surrounding Halifax Harbor, Nova Scotia. *Mar. pollut. Bull.* 81 (1), 69–79. doi: 10.1016/j.marpolbul.2014.02.018

Mehinto, A. C., Coffin, S., Koelmans, A. A., Brander, S. M., Wagner, M., Thornton Hampton, L. M., et al. (2022). Risk-based management framework for microplastics in aquatic ecosystems. *Micro Nanoplastics* 2, 17. doi: 10.1186/s43591-022-00033-3

Nair, H. T., and Perumal, S. (2022). Trophic transfer and accumulation of microplastics in freshwater ecosystem: risk to food security and human health. *Internat. J. Ecol.* 2022, 1234078. doi: 10.1155/2022/1234078

Nelms, S. E., Galloway, T. S., Godley, B. J., Jarvis, D. S., and Lindeque, P. K. (2018). Investigating MP trophic transfer in marine top predators. *Environ. pollut.* 238, 999–1007. doi: 10.1016/j.envpol.2018.02.016

Neves, D., Sobral, P., Ferreira, J. L., and Pereira, T. (2015). Ingestion of MPs by commercial fish off the Portuguese coast. *Mar. pollut. Bull.* 101 (1), 119–126. doi: 10.1016/j.marpolbul.2015.11.008

Nielsen, B. L. H., Gréve, H. V. S., and Hansen, B. W. (2020). Cultivation success and fatty acid composition of the tropical copepods *Apocyclops royi* and *Pseudodiaptomus annandalei* fed on monospecific diets with varying PUFA profiles. *Aquaculture Res.* 52 (3), 1127–1138. doi: 10.1111/are.14970

Oaten, A., and Murdoch., W. W. (1975). Functional response and stability in predator-prey systems. *Am. Nat.* 109, 289–298. Available at: http://www.jstor.org/stable/2459696.

Paffenhöfer, G.-A., and Van Sant, K. B. (1985). The feeding response of a marine planktonic copepod to quantity and quality of particles. *Mar. Ecol. Prog. Ser.* 27, 55–65. doi: 10.3354/meps027055

Pan, D., Su, F., Liu, H., Liu, C., Umar, A., Castañeda, L., et al. (2021). Research progress on catalytic pyrolysis and reuse of waste plastics and petroleum sludge. *ES Mat Manuf* 11, 3–15. doi: 10.30919/esmm5f415

Qiao, R., Lu, K., Deng., Y., Ren, H., and Zhang, Y. (2019). Combined effects of polystyrene microplastics and natural organic matter on the accumulation and toxicity of copper in zebrafish. *Sci. Total Environ.* .682, 128–137. doi: 10.1016/j.scitotenv.2019.05.163

Rajan, K., Khudsar, F. A., and Kumar, R. (2023). Spatio-temporal patterns of microplastic contamination in surface waters of Hooghly River Estuary: Causes and consequences. *Regional Stud. Mar. Sci.* 65, 103111. doi: 10.1016/j.rsma.2023.103111

Rao, T. R., and Kumar, R. (2002). Patterns of prey selectivity in the cyclopoid copepod *Mesocyclops thermocyclopoides*. *Aquat. Ecol.* 36, 411–424. doi: 10.1023/A:1016509016852

Richon, C., Gorgues, T., Paul-Pont, I., and Maes, C. (2022). Zooplankton exposure to MPs at global scale: Influence of vertical distribution and seasonality. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.947309

Roch, S., Friedrich, C., and Brinker, A. (2020). Uptake routes of MPs in fishes: Practical and theoretical approaches to test existing theories. *Sci. Rep.* 10, 3896. doi: 10.1038/s41598-020-60630-1

Rochman, C. M., Brookson, C., Bikker, J., Djuric, N., Earn, A., Bucci, K., et al. (2019). Rethinking microplastics as a diverse contaminant suite. *Environ. Toxicol. Chem.* 38, 703–711. doi: 10.1002/etc.4371

Rochman, C., Hoh, E., Kurobe, T., and Teh, S. J. (2013). Ingested plastic transfers hazardous chemicals to fish and induces hepatic stress. *Sci. Rep.* 3, 3263. doi: 10.1038/srep03263

Sarma, S. S. S., and Nandini, S. (2002). Studies on functional response and prey selection using zooplankton in the anostracan *Chirocephalus diaphanus* Prevos. *Hydrobiologia* 486, 169–174. doi: 10.1023/A:1021398718074

Setälä, O., Fleming-Lehtinen, V., and Lehtiniemi, M. (2014). Ingestion and transfer of MPs in the planktonic food web. *Environ. pollut.* 185, 77–83. doi: 10.1016/j.envpol.2013.10.013

Smith, B. E., and Smith, L. A. (2020). Multispecies functional responses reveal reduced predation at high prey densities and varied responses among and within trophic groups. *Fish Fisheries* 21, 891–905. doi: 10.1111/faf.12468

Solomon, M. E. (1949). The natural control of animal populations. *J. Anim. Ecol.* 18, 1–35. doi: 10.2307/1578

Svetlichny, L., İşinibilir, M., Mykitchak, T., Eryalçın, K. M., Türkeri, E. E., and Yüksel, E. (2021). Microplastic consumption and physiological response in *Acartia clausi* and *Centropages typicus*: possible roles of feeding mechanisms. *Regional Stud. Mar. Sci.* 43, 101650. doi: 10.1016/j.rsma.2021.101650

Ter Halle, A., Ladirat, L., Martignac., M., Mingotaud, A. F., Boyron, O., and Perez, E. (2017). To what extent are MPs from the open ocean weathered? *Environ. pollut.* 227, 167–174. doi: 10.1016/j.envpol.2017.04.051

Traboni, C., Sarno, D., Ribera d'Alcalà, M., and Mazzocchi, M. G. (2023). Microplastics in the menu of Mediterranean zooplankton: Insights from the feeding response of the calanoid copepod *Centropages typicus*. *Mar. Ecol.* 00, e12760. doi: 10.1111/maec.12760

Tseng, L. C., Kumar, R., Chen, Q. C., and Hwang, J. S. (2011). Summer distribution of *noctiluca scintillans* and mesozooplankton in the Western and Southern East China Sea prior to the three gorges dam operation. *Hydrobiologia* 666 (1), 239–256. doi: 10.1007/s10750-010-0587-y

Turner, J., and Tester, P. (1989). Zooplankton feeding ecology: copepod grazing during an expatriate red tide. 359–374. doi: 10.1007/978-3-642-75280-3_21

Ugwu, K., Herrera, A., and Gomez, M. (2021). Microplastics in marine biota: A review. *Mari Pol. Bul.* 169, 112540. doi: 10.1016/j.marpolbul.2021.112540

Vanderploeg, H. A., Paffenhöfer, G. A., and Liebig, J. R. (1990). Concentration-variable interactions between calanoid copepods and particles of different food quality: observations and hypotheses. *Behav. Mech. Food Selection* 20, 595–613. ISBN: 978-3-642-75120-2.

von Moos, N., Burkhardt-Holm, P., and Köhler, A. (2012). Uptake and effects of microplastics on cells and tissue of the blue mussel *Mytilus edulis* L. after an experimental exposure. *Environ. Sci. Technol.* 46 (20), 11327–11335. doi: 10.1021/es302332w

Walne, P. R. (1970). Studies on the food value of nineteen genera ofalgae to juvenile bivalves of the genera Ostrea, Crassostrea, Mercenaria,and Mytilis. *Fishery Investigations* 26, 1–62.

White, J. R., and Dagg, M. J. (1989). Effects of suspended sediments on egg production of the calanoid copepod Acartia tonsa. *Mar. Biol.* 102, 315–319. doi: 10.1007/BF00428483

Woods, M. N., Hong, T. J., Baughman, D., Andrews, G., Fields, D. M., and Matrai, P. A. (2020). Accumulation and effects of microplastic fibers in American lobster larvae (*Homarus americanus*). *Marine Pollu. Bulletin* 157, 111280.

Xu, J., Rodríguez-Torres, R., Rist, S., Nielsen, T. G., Hartmann, N. B., Brun, P., et al. (2022). Unpalatable plastic: efficient taste discrimination of MPs in planktonic copepods. *Environ. Sci. Technol.* 56 (10), 6455–6465. doi: 10.1021/acs.est.2c00322

Zhang, F., Wang, X., Xu, J., Zhu, L., Peng, G., Xu, P., et al. (2019). Food-web transfer of MPs between wild caught fish and crustaceans in East China Sea. *Mar. pollut. Bull.* 146, 173–182. doi: 10.1016/j.marpolbul.2019.05.061

Zimmermann, L., Dierkes, G., Ternes, T. A., and Völker, C. (2019). Wagner M. Benchmarking the in *vitro* toxicity and chemical composition of plastic consumer products. *Environ. Sci. Technol.* 53 (19), 11467–11477. doi: 10.1021/acs.est.9b02293

View publication stats

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-18637-7

Modeling and simulation of FIR filter using distributed arithmetic algorithm on FPGA

Amrita Rai[1] · Ajay Roy[2] · Shamimul Qamar[3] · Abdulelah G. F. Saif[3] · Magdi Mohammad Hamoda[3] · Abdul Azeem[4] · Salman Arafath Mohammed[5]

Received: 17 April 2022 / Revised: 22 December 2023 / Accepted: 12 February 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
In many industries and telecommunication system there is a need for digital signal processing for fast transfer of data between two points or devices with low power consumption and considerable hardware resources (circuit size and speed). Finite Impulse Response (FIR) filters play an important role in many signal processing applications and telecommunication systems. This paper propose the design and implementation of 4-bit FIR Filter using Distributed Arithmetic (DA) Algorithms, which substitute, multiply and accumulate operation with series of Look Up Table (LUT). The proposed FIR filter is implemented in high-density field programmable logic devices (FPGAs) and designed using very high-speed integrated circuit hardware description language (VHDL) and verified using Xilinx ISE 14.7 tool and simulator. The proposed, modified and optimized DA provided the multiplication and accumulation free calculation of inner product data of FIR filter and this consecutively reduces the size and power dissipation of circuit. DA is one of the methods to implement FIR filters that impact the storage resource and the calculating speed, which make the memory size smaller and the operation speed faster. The simulated proposed structure required nearly 40% less cells, 35% less LUT pairs and 4% less power consumption with existing structure.

Keywords Telecommunication · Distributed arithmetic algorithms · Look up table (LUT) · Finite impulse response (FIR) filter · FPGA Artix-7™

1 Introduction

Finite impulse response (FIR) filters are used in video and communication systems where high performance in terms of speed, area, and power are required. Digital filters are used to modify the features of signals in the time and frequency domains. The design of FIR filters has focused on multiplier-based methods such as multiply-and-accumulate (MAC) [1, 8]. Recently, high-speed and high-order programmable FIR filters have been frequently used

✉ Shamimul Qamar
sqamar@kku.edu.sa; drsqamar@rediffmail.com

Extended author information available on the last page of the article

to perform adaptive pulse shaping and signal equalization on received data in real time, such as ghost cancellation and channel equalization [2]. Thus, an effective very large scale integration (VLSI) architecture for high-speed programmable FIR filters is required [3]. MAC blocks, which are key components of FIR filters and many other circuits, have been implemented in digital signal processing (DSP) primarily using multiplier-based architectures. Fast parallel-filter architectures have been studied in depth [4–8]. FIR filters are the fundamental building blocks of many signal processing and communication applications [5].

The MAC method of FIR filter implementation is expensive to implement in a field-programmable gate array (FPGA) owing to logic complexity and resource usage [6]. To resolve this problem, many researchers have used the distributed arithmetic (DA) algorithm in the FIR filter design process [9, 15]. In the MAC method, any arithmetic operation is implemented by first multiplying and then adding; however, the DA algorithm converts this by shifting and adding. The DA algorithm has been further improved by removing the configuration problem in the coefficient of the FIR filter and calculating the speed and memory size [7, 10].

A computerized channel is a framework that performs numerical operations on a discrete-time signal that has been examined to change or modify a portion of the signal in either the time or recurrence space. It has a basic-to-advanced converter at the front end, followed by a chip and several peripheral components, such as memory, to store data and channel coefficients [8]. Mechanical advancements have improved the various zones of DSP, one of which is the structure of all-around arranged calculations to ascertain the discrete Fourier transform [9]. The introduction of programmable digital signal processors (PDSPs) in the late 1970s was another advancement in this field that provided the option to perform multiplication and addition in one clock cycle. Progressively improved capacities, such as memory banks, gliding point multipliers, or zero-overhead interfaces to analog-to-digital converters and digital-to-analog converters, are incorporated into cutting-edge PDSPs [10–15].

The use of FIR channels in video and communication systems depends on their performance, which necessarily depends on the speed, size, and power consumption. Fundamentally, computerized channels are used to change the normal sign in time and recurrence areas and have been considered essential advanced signal processing. In DSP, the plan strategies are essentially engaged in multiplier-based designs to execute the increase and accumulate obstructs that establish the focal component in FIR channels and a few capacities [11]. In this study, complicated and time-consuming MAC techniques were replaced with shifting and adding. To achieve the best configuration of the FIR filter, we pre-store the coefficients in a look-up table (LUT) to reduce the multiplication time. The design improves significantly when compared with regular FPGA acknowledgment, and it can also be applied to conduct high-pass, low-pass, and band-stop channels by changing the request and LUT coefficients [12, 13]. The suggested approach and present method were designed and implemented using Verilog HDL coding and a Xilinx synthesizer. The Xilinx ISE design software offers a full, multi-platform design environment that is easily adaptable to unique design requirements. This is a suitable environment for designing a system-on-a-chip (SOC). Figure 1 shows the solutions that Xilinx software offers for every stage of the FIR filter design using the DA algorithm on either an FPGA or complex programmable logic device (CPLD) implementation board. The figure explains each stage of the design flow required for software and hardware implementation.

This paper primarily focuses on the design and implementation of a finite impulse response using the DA algorithm with low power consumption and high speed using an FPGA. The

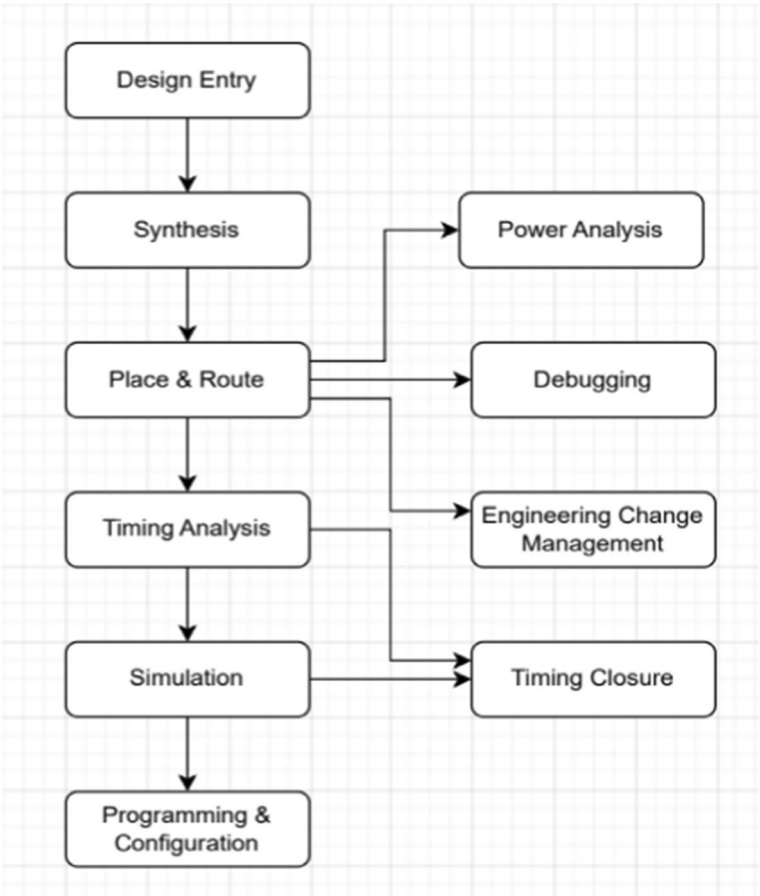


Fig. 1 Flow diagram of the proposed design in software and hardware

principal objective of this study was a detailed investigation of the equipment usage of FIR channels using the DA Algorithm. Simulations were performed using software and an FPGA kit. A Nexys 4 DDR board was used as the hardware. The software used was Xilinx ISE Design Suite 14.7.

2 Basic introduction of the FIR filter and DA algorithm

High performance in terms of speed, size, and power consumption is required owing to the extensive use of FIR filters in video and communication systems [14, 15]. Digital filters are used in DSP to change the properties of signals in the time and frequency domains [16]. Generally, an FIR filter is mathematically described as the sum of the convolution of the input and impulse responses, as shown in Eq. 1.

$$Y(n) = h(n) * X(n) \tag{1}$$

$$Y(n) = \sum_{m=0}^{N-1} h(m)x(n-m) \tag{2}$$

where $X(n)$ is the input sequence of the filter; $h(n)$ is the impulse response, and N is the number of filter taps. Equation (2) shows that the output of an FIR filter is the result of multiplying the impulse response by the input sequence, which is a long series of multiplication operations. Many researchers have used multiplier-less structures or designs for FIR filter implementation because multipliers require a large area. DA is one of the best-known methods for implementing FIR filters. DA solves the computation of the inner product equation when the coefficients are known, as in FIR filters [17]. Equation 3 expresses the FIR channels using numerical articulation at any instant of time n and the output of the M-tap FIR filter in simplified form.

$$y(n) = \sum_{m=0}^{N-1} B_m X_m \tag{3}$$

where m represents the FIR channel taps; X_m represents the tap coefficients with class M, that is, the reaction of unit motivation; $x(n-m)$ represents the information signal postponing K taps; $x(n)$ and $Y(n)$ represent the input and output, respectively. The fundamental structure of a direct-stage FIR channel is shown in Fig. 2 [18]. Similarly, the key aspect of FIR channel planning is the count of K-times convolution, which requires a significant increase in storage memory for each stage of data execution. Therefore, we utilize an improved DA calculation to address this problem. As DA is an exceptionally proficient arrangement, particularly appropriate for LUT-based FPGA designs, several analysts have used it extensively to actualize FIR filters in FPGAs [19].

An LUT in the DA calculation can enhance the convolution task, which is required for implementing FIR filters. The DA calculation comprises sequential DA calculations, parallel DA calculations and joined arrangements and parallel DA calculations for different filter measurement characteristics. Sequential dispersed calculations generally have a straightforward structure and are less involved as sets; however, their speed is low because of the length of the information [20]. Parallel-circulated calculations have a slick structure and are generally utilized in high-speed events. However, the parallel conveyed calculation will result in a higher utilization of assets. Regardless of the type of dispersed calculation, read-only memory (ROM) is selected for the LUT. Analyses demonstrate that with an increase in the number of channel requests, the quantity of

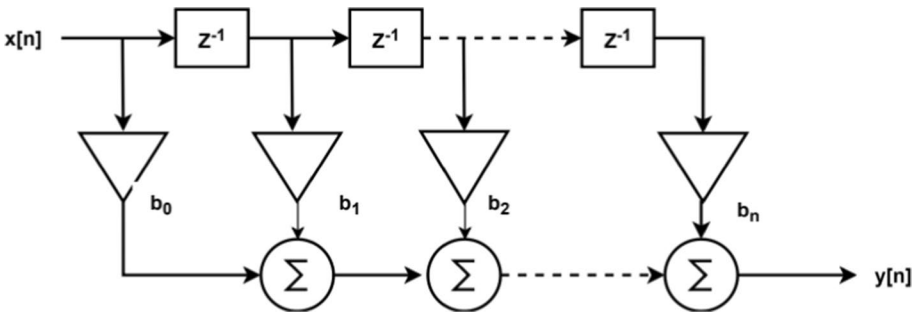


Fig. 2 Basic parallel structure of an FIR filter

ROM dependent on the two controls also expands. Thus, we enhance the LUT structure of the consolidated arrangement and parallel DA calculation to improve both the handling rate and sparing rationale assets [21].

DA algorithm DA is a well-known method for implementing FIR filters. The essential concept is to replace addition and multiplication with an LUT and shifter–accumulator pair. DA depends on the known channel coefficients; thus, duplicating c[n]x[n] terms into the derived equation does not change, but it is consistent and repeated. This is a significant contrast and is essential for a DA plan. This is an excellent strategy for reducing the size of equal duplicate collection equipment, which is appropriate for FPGA boards [7, 10, 14].

The output of the linear time-invariant system is given by Eq. (4).

$$AY[n] = A_1X_1 + A_2X_2 + A_3X_3 + ... + A_mX_m \tag{4}$$

$$Y[n] = \sum_{m=1}^{M} AmXm \tag{5}$$

where A_m is a fixed factor, X_m represents the N-bit input data, and $|X_m|<1$.

X_m can be expressed as Eq. (6) using the binary complement:

$$X_m = -x_{m0} + \sum_{n=1}^{N-1} x_{mn} * 2^{-n} \tag{6}$$

where x_{mn} is 0 or 1, x_{m0} is the sign bit, and x_{mN-1} is the least significant bit (LSB). Thus, $Y[n]$ can be expressed by Eq. (7).

$$\begin{aligned} Y[n] &= \sum_{m=1}^{M} A_m \left(\sum_{n=1}^{N-1} x_{mn} 2^{-n} - x_{m0} \right) \\ Y[n] &= \sum_{n=1}^{N-1} \sum_{m=1}^{M} A_m 2^{-n} x_{mn} + \sum_{m=1}^{M} A_m(-x_{m0}) \end{aligned} \tag{7}$$

$$Y[n] = \sum_{m=1}^{M} [(A_m * x_{m1})2^{-1} + (A_m * x_{m2})2^{-2} + (A_m * x_{m(N-1)})2^{N-1}] + \sum_{m=1}^{M} A_m(-x_{m0}) \tag{8}$$

$$\begin{aligned} Y[n] = {} & [(A_1 * x_{11})2^{-1} + (A_1 * x_{12})2^{-2} + ... + (A_1 * x_{(N-1)})2^{N-1}] + [(A_2 * x_{21})2^{-1} + (A_2 * x_{22})2^{-2} \\ & + ... + (A_2 * x_{2(N-1)})2^{N-1}] \\ & + ... + [(A_M * x_{M1})2^{-1} + (A_M * x_{M2})2^{-2} + ... \\ & + (A_M * x_{M(N-1)})2^{N-1}] - [(A_1 * x_{10}) + (A_2 * x_{20}) + ... + (A_M * x_{M0})] \end{aligned} \tag{9}$$

$$\begin{aligned} Y[n] = {} & -[(A_1 * x_{10}) + (A_2 * x_{20}) + ... + (A_M * x_{M0})] \\ & + [(A_1 * x_{11}) + (A_2 * x_{21}) + \ldots ... + (A_M * x_{M1})]2^{-1} + [(A_1 * x_{12}) + (A_2 * x_{22}) \\ & + \ldots . + (A_M * x_{M2})]2^{-2} + ... + [(A_1 * x_{(N-1)}) + (A_2 * x_{2(N-1)}) + ... + (A_M * x_{M(N-1)})]2^{N-1} \end{aligned} \tag{10}$$

$$Y[n] = -\sum_{m=1}^{M} A_m(-x_{m0}) + \sum_{n=1}^{N-1} (A_1x_{1n} + A_2X_{2n} + \ldots .. + A_MX_{Mn})2^{-n} \tag{11}$$

$$Y[n] = \sum_{n=1}^{N-1} \sum_{m=1}^{M} A_m 2^{-n} x_{mn} + \sum_{m=1}^{M} A_m(-x_{m0}) \tag{12}$$

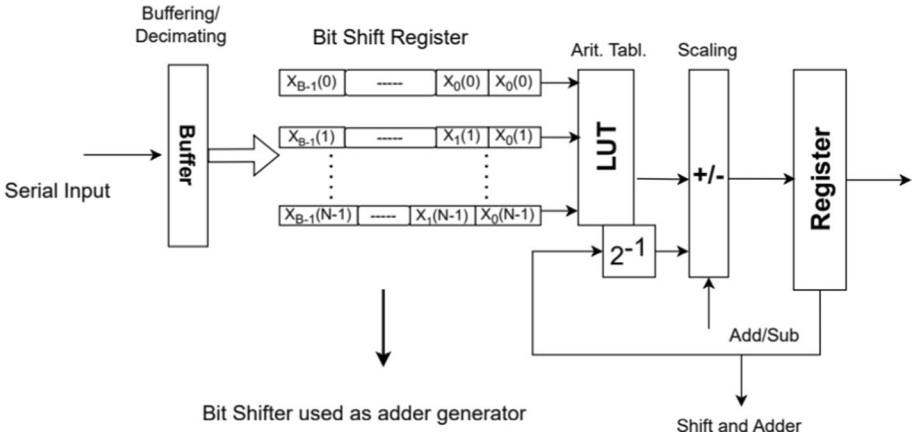


Fig. 3 Distributed Arithmetic implementation of a FIR filter

The above equations are used to configure the 4-bit FIR filter using the DA algorithm, and a basic diagram of the DA-based FIR filter implementation is shown in Fig. 3. While executing DA, it is important to store the contribution of the coefficient length in the buffer stage. When it is complete, the LSB of all coefficients carries the location to the LUT. That is, the second word LUT is pre-modified to acknowledge an N-bit address, where N is the coefficient length. For execution using the FPGA, rather than moving each moderate word by exponential power, which requires a costly barrel shifter, the moderated word itself is shifted towards every path of the consecutive bits to one side using the proposed algorithm.

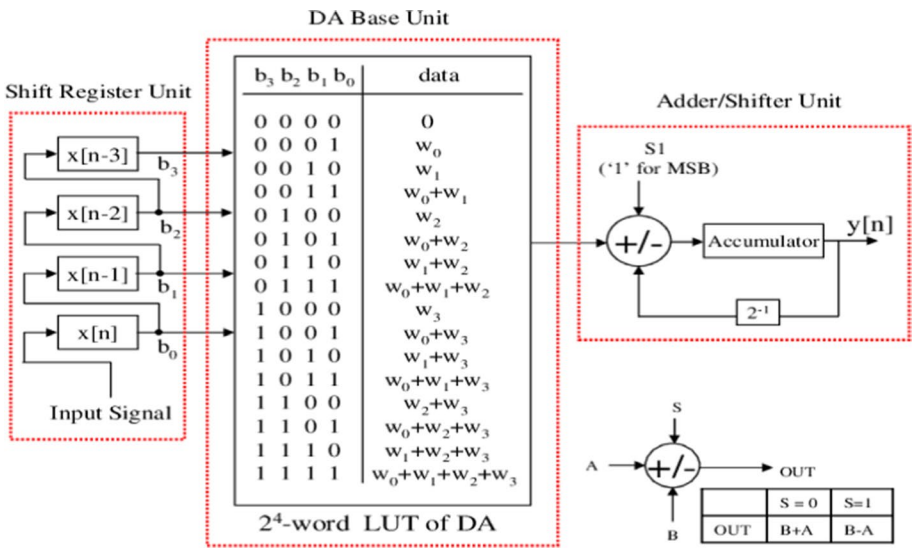


Fig. 4 Actual Hardware Implementation of 4 tap FIR filter with DA algorithm

3 Circuit diagram of the 4-tap FIR filter with the DA algorithm

The conventional LUT-based DA architecture for a 4-tap FIR filter is shown in Fig. 4. The design index and parameters were selected for simplicity in implementation and simulation, and the simulation parameters of the design filter were obtained at a sampling frequency of 25 MHz; the input and output data were 8 bits, and the filter coefficients were 2, -4, 1, 2. All hardware requirements of the 4-tap FIR filter with the LUT-based DA architecture are shown in Fig. 4.

From Eqs. (1)–(10), we conclude that

$$\sum_{m=1}^{M} A_m 2^{-n} \text{ has } 2^M \text{ Possible Values}$$
$$\sum_{m=1}^{M} A_m\left(-x_{m0}\right) \text{ has } 2^M \text{ Possible Values}$$

Along with the sign bit, m can be stored in a 2×2^{M} ROM storage. If the number of taps is 4, then the total memory required is 2^4, indicating a 16-word length memory. The ROM size increases exponentially with each input address line. If the address line is K with a value of 16, this implies that 2^{16} (i.e., 64 K) of ROM is required for normal implementation. However, the use of the DA algorithm and LUT can be minimized. Up to 80% of the area can be saved using DSP hardware designs when DA is executed in FPGAs, and we can exploit the memory in FPGAs to implement the MAC activity and LUT [10].

4 Simulation and implementation of the proposed system with the FPGA kit

The simulations were performed using the Xilinx Tools ISE Design Suite 14.7. Xilinx Tools is a suite of programming instruments for creating advanced circuits in FPGAs or CPLDs. In this study, an FPGA trainer kit was used to verify the design. A red light indicated that the FPGA kit is on, and a green light indicated that the FPGA kit was programmed. It had 16 light-emitting diodes (LEDs) and 16 slide switches. The input voltage was 4.5–5.5 V with a current of 1 A. The total power required was approximately 5 W.

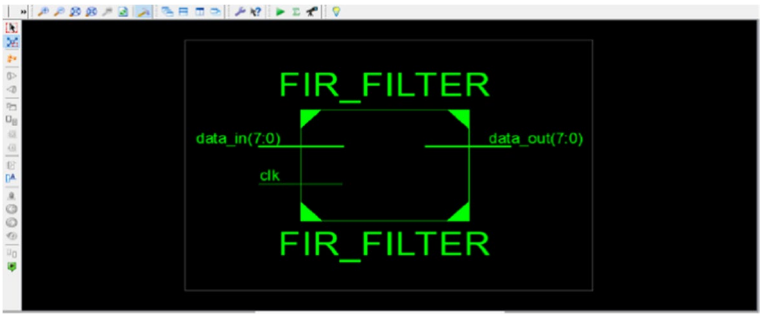


Fig. 5 Block representation of synthesis results of 4 tap fir filter

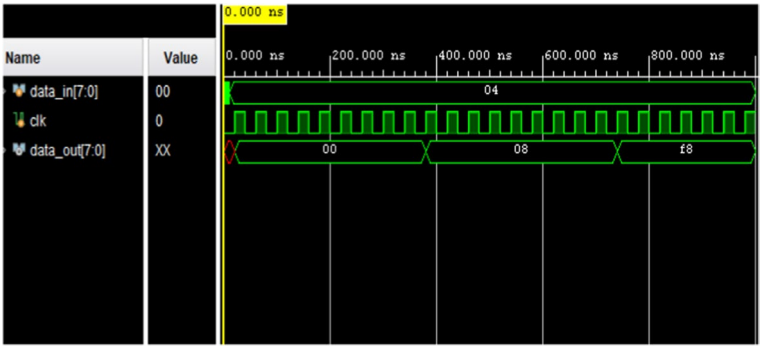


Fig. 6 Simulation results of 4 tap finite impulse response filter

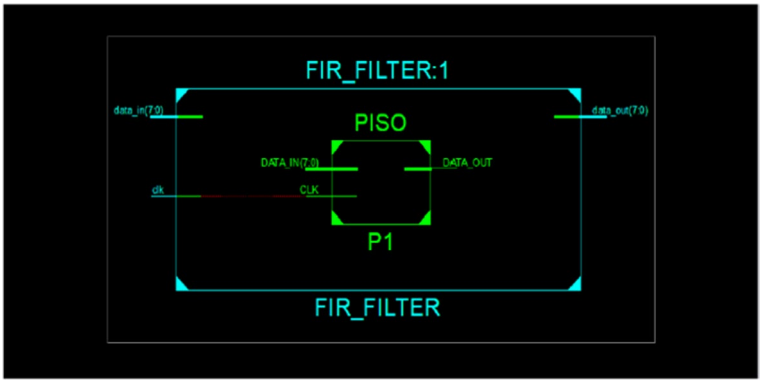


Fig. 7 RTL view of proposed 4-tap FIR Filter

The proposed setup was a 4-tap FIR filter with an 8-bit input and output. Hence, a width of 8 bits was used. One parallel-in serial-out (PISO) register (P1) and three serial-in serial-out (SISO) registers (S1, S2, and S3) were initialized. The coefficients 2, -4, 1, and 2 were considered for the 4-tap FIR filter. With the 4-bit input applied to the FIR filter, an 8-bit output with coefficients of -8, -4, and 4 was generated. The DA algorithm relies on the fact that all possible combinations are stored to avoid multiplication. Therefore, we designed a 16-address LUT containing all the possible combinations. The main function of the accumulator is to store previous data and add them to new data. Hence, the size of the accumulator was double the input size, i.e., 16 bits. Figure 5 shows the block representation of the designed FIR filter using the DA algorithm, and Fig. 6 depicts the 4-tap FIR filter output–input simulation of the proposed improved design based on the DA algorithm.

To perform synthesis, we first selected the main function as UUT-FIR_FILTER. Additionally, we examined the RTL and technology schematics, as shown in Fig. 7. To perform the simulation, we first selected the test bench of the FIR filter, TB_FIR_FILTER. Figure 8 shows a detailed RTL view of the FIR filter design using the DA algorithm.

Table 1 lists the cells, number of slice LUTs (SLUTs), and LUT flip-flop pairs used. Compared with the existing DA-based FIR filter structure, the proposed structure used

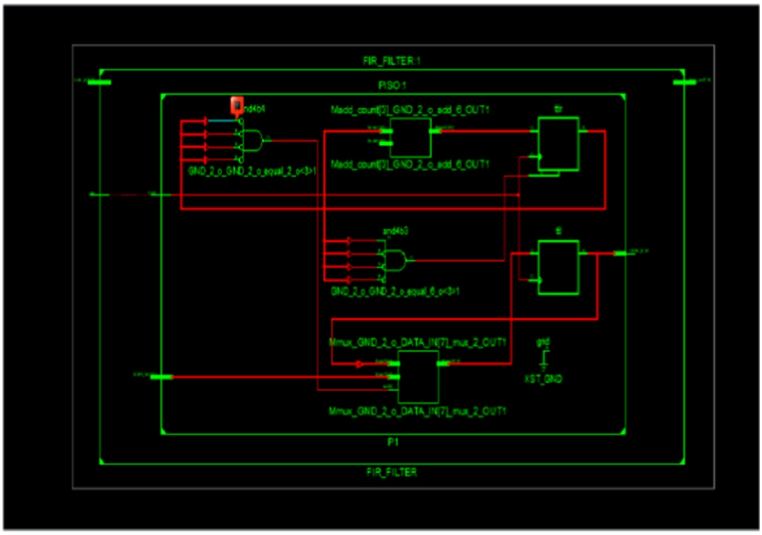


Fig. 8 Detail synthesis results of 4 tap FIR filter

Table 1 Parameter comparison for the existing and modified DA-based FIR filter architectures

Parameters	Existing DA-based FIR Filter Structure	Modified DA-based FIR Filter Structure	Improvement
Cells used	606	357	42% fewer cells
Slice LUTs	567 (2%)	356 (1%)	1% of total LUTs
LUT–flip-flops pairs	656	401	40% fewer flip-flops
Power	387 mW	379 mW	2% less power
Minimum time	2.375 ns	2.523 ns	7% more time

approximately 42% fewer cells, 40% fewer LUT flip-flop pairs, and 2% less power. In addition, the proposed structure required the area-delay product to be lowered to 37% using the improved DA-based FIR filter. The existing structure used more power than the modified structure because parallel processing requires more power owing to more computations; however, although the area decreased by 40%, the power reduction was insufficient. However, the area decreased significantly when the partial product generation block was reused L times.

5 Hardware implementation of the proposed 4-tap FIR filter using the FPGA kit

An FPGA is an incorporated circuit intended to be arranged by a researcher or student after assembly through a field-programmable board. The FPGA arrangement is commonly determined using an equipment hardware description language (HDL), similar to that utilized for an application-specific integrated circuit (ASIC). Circuit graphs have

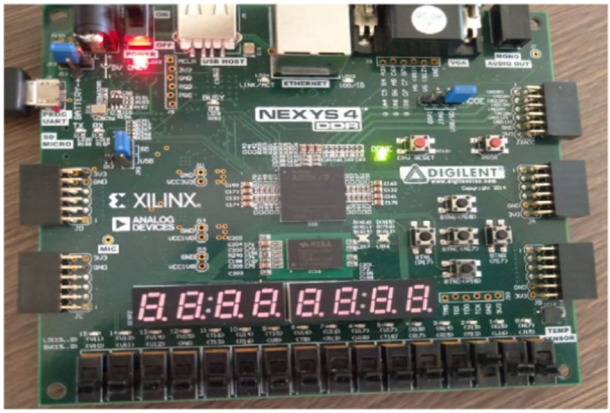


Fig. 9 A Nexys 4 DDR

recently been used to indicate the setup, as they are for ASICs; however, this is progressively uncommon. A Spartan FPGA from Xilinx contains various programmable gate arrays and reconfigurable interconnects that enable the gates to be wired together, similar to other rational entryways that can be wired in various designs. Rationale gate arrays can be designed to perform complex combinational capacities or only basic rationale functions, such as AND, OR, and XOR. In many FPGAs, gate arrays also incorporate memory components, which may be a basic flip-flop or a progressively complete set of memory. The Nexys DDS 4 kit, shown in Fig. 9, operates in two modes: the FPGA is controlled or customized. The green light indicates that the FPGA has some programs, and the red light indicates that the FPGA is powered up. The FPGA can be reset by holding the switch program and reset buttons together.

Whatever was tested in the ISE Design Suite could also be implemented using a hardware kit. Similar to the simulation, the coefficients of the proposed filter were 2, -4, 1, and 2. If the input is 4, the outputs will be 8, -8, -4, and 4 is implemented in the FPGA kit. The first output of the one-tap FIR filter is shown in Fig. 10.

The analysis of the algorithm to be implemented using the FPGA kit for different outputs of the FIR filter is shown in Figs. 11, 12, and 13 for all four combinations of outputs. Toggle-switch buttons in FPGA kits were programmed as input signals, and sequential green LEDs were assigned as the output of the proposed FIR 4-tap filters. Sixteen switch buttons were connected to the programmable FPGA IC within the kit, and the 16 green LEDs just above the toggle buttons indicated a 16-bit output. From the left side of the

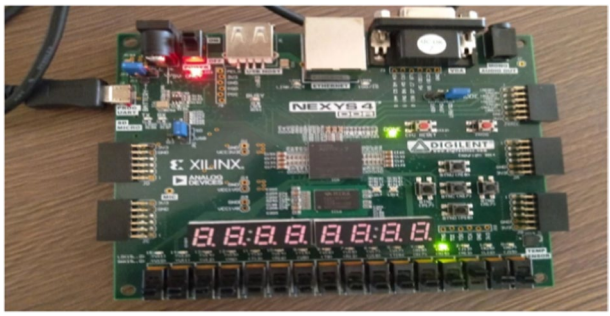


Fig. 10 First output of 4 tap FIR filter

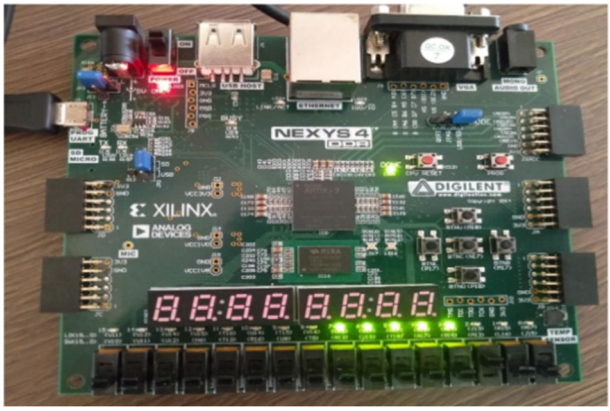


Fig. 11 Second output of 4 tap Fir filter

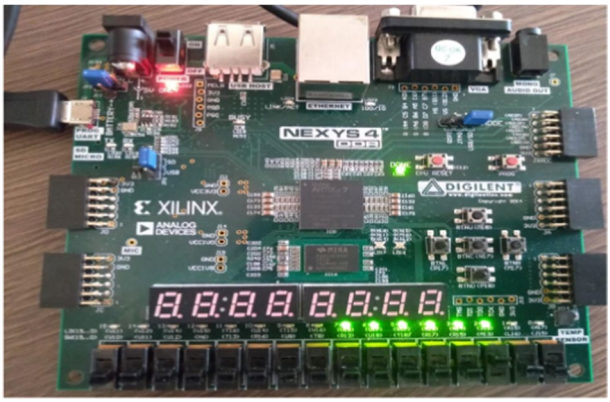


Fig.12 Third output of 4 tap Fir filter

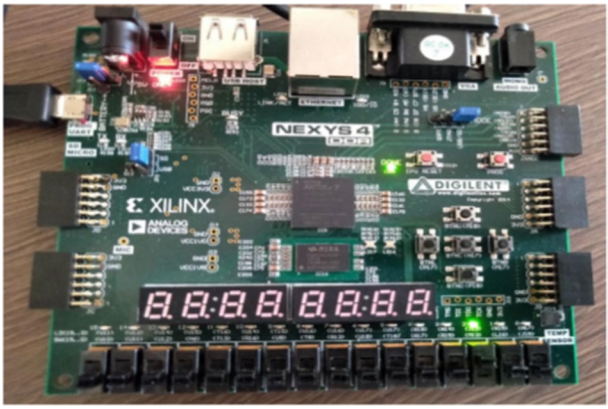


Fig. 13 Fourth output of 4 tap Fir filter

kit, the first four toggle buttons and green LED were used as inputs and outputs, respectively. All the toggle buttons were programmed and set according to the assigned input. Figures 10, 11, 12 and 13 show the different combinations of outputs according to the input and filter coefficients and the above-calculated value. As shown in Fig. 10, LED 4 was on and the other three were off, indicating that the output was 1000, i.e., 8. Similarly, Figs. 11, 12, and 13 represent the different sets of on and off LEDs.

The filter input/output in the waveform used a hexadecimal representation. Compared with a MATLAB simulation, the error of hardware simulation was $\leq \pm 1\%$. The circuit structure of the FIR filter operated well, and the simulation showed the results.

6 Conclusion

After reviewing all types of DA structures for FIR filter implementation, the proposed effective FIR filter implementation was based on modified and optimized DA algorithms. This model proposes another calculation for FIR advanced channel unions for many fixed coefficients to reduce the number of adders and limit the wire delay. In this study, the filter structure was implemented for a 4-tap FIR filter and applied for more taps, with the use of split LUT techniques of DA for multiplication and accumulation of filter coefficients. The design was verified using an Nexys DDS 4 FPGA kit. The simulated and implemented results indicated that nearly 40% fewer cells were required, a smaller area was used, the LUT flip-flop requirement decreased by 35%, and power consumption decreased by 4% compared with the existing structure, as shown in Table 1. For the different signal processing and telecommunication application the various parameter and order of filter can be modified accordingly.

Acknowledgements The authors would like their gratefulness to the Deanship of Scientific Research, King Khalid University, Abha, Saudi Arabia, to provide administrative, financial, and technical support under Grant Number GRP2/ 170/1444.

Declarations

Research involving human participants and/or animals This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest The authors have no conflict of interest.

References

1. Gaurav P (2020) FIR filter design and implementation using FPGAs. https://www.linkedin.com/pulse/fir-filter-design-implementation-using-fpgas-gaurav-pahuja
2. Zheng J and Wei Z (2018) FIR filter design based on FPGA, 2018 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Changsha, China, pp. 36–40. https://doi.org/10.1109/ICMTMA.2018.00016
3. Bhavya V, Pallavi JP, Abhilash SV (2021) Implementation of Fir filter using distributed arithmetic and distributed arithmetic off-set binary coding algorithms. Int J Adv Res Sci Commun Technol (IJARSCT) 7(2). https://doi.org/10.48175/IJARSCT-1770
4. Ariprasath S, Santhi C (2017) Transpose form FIR filter design for fixed and reconfigurable coefficients. Int Res J Eng Technol (IRJET) 4(3):1859

5. Rai A (2023) An optimization of low power 4-bit PAL FIR filter using adiabatic techniques. Sādhanā 48:84. https://doi.org/10.1007/s12046-023-02132-0
6. Bir P, Karatangi SV, Rai A (2020) Design and implementation of an elastic processor with hyper-threading technology and virtualization for elastic server models. J Supercomput 76(9):7394–7415. https://doi.org/10.1007/s11227-020-03174-5
7. Amita N, Vigneswarn T, Ashwani KR, Arvind D (2015) An efficient 256-tap parallel FIR digital filter implementation using distributed arithmetic architecture. Procedia Comput Sci 54:605–611. https://doi.org/10.1016/j.procs.2015.06.070. ISSN 1877–0509
8. Ifeachor E, Jervis B (2002) Digital signal processing: a practical approach, 2nd edn. Prentice Hall, Harlow, pp 690–707
9. Haj Al, Ali M (2003) Fast discrete wavelet transformation using FPGAs and distributed arithmetic. Int J Appl Sci Eng 1:160–171. https://doi.org/10.6703/IJASE.2003.1(2).160
10. Nian LQ, Hou SY, Cui SY (2010) Application of distributed FIR filter based on FPGA in the analyzing of ECG signal. 2010 International Conference on Intelligent System Design and Engineering Application, Changsha, China, pp. 335–338. https://doi.org/10.1109/ISDEA.2010.265
11. Rai SN, Shree BSP, Meghana YP (2018) Design and implementation of 16 tap FIR filter for DSP applications. Proceding of Second International Conference Advance in Electronics Computer and Communication, Bengaluru, India, pp 9–10
12. Nayak S, Rai A (2021) Synthesis and analysis of optimal order Butterworth filter for denoising ECG signal on FPGA. In Information Management and Machine Intelligence: Proceedings of ICIMMI 2019. Springer, Singapore, pp 359–369
13. Zhang Y, Zhu Y, Han K, Wang J, Hu J (2021) A high-accuracy stochastic FIR filter with adaptive scaling algorithm and antithetic variables method. Electronics 2021(10):1937. https://doi.org/10.3390/electronics10161937
14. Radhika SG, Weijia S, Qiang L (2002) A faster distributed arithmetic architecture for FPGAs. Proceedings of the 2002 ACM/SIGDA Tenth international symposium on Field-programmable gate arrays, pp 31–39. https://doi.org/10.1145/503048.503054
15. Sinhmar P, Rai A (2012) Modeling and Simulation of FIR Filters and Comparison of Frequency Response in Time Domain and Frequency Domain for UMTS. International Journal of Electronics Communication and Computer Engineering. http://www.ijecce.org › files › IJECCE- 199pa 3(1):216–221
16. Naveen SN, Kiran G (2015) An efficient reconfigurable FIR digital filter using modified distribute arithmetic technique. Int J Emerg Technol Adv Eng 5(6). Website: www.ijetae.com. https://arxiv.org/abs/1704.08526
17. Muktesh KO et al (2022) Cuckoo search constrained gamma masking for MRI image detail enhancement. Traitement du Sig 39(4). https://doi.org/10.18280/ts.390433
18. Vinay R, Vijayakumar TSVS, Saini LM, Singh B (2020) Power efficient FIR filter architecture using distributed arithmetic algorithm. First IEEE International Conference on Measurement, Instrumentation, Control and Automation , Kurukshetra, India, 2020, pp 1–4.https://doi.org/10.1109/ICMICA48462.2020.9242720
19. Sharma D, Rai A, Debbarma S, Prakash O, Ojha MK, Nath V (2023) Design and optimization of 4-bit array multiplier with adiabatic logic using 65 nm CMOS technologies. IETE J Res 1–14. https://doi.org/10.1080/03772063.2023.2204857
20. Sangyun H, Gunhee H, Sungho K, Jaeseok K (2004) New distributed arithmetic algorithm for low-power FIR filter implementation. IEEE Signal Process Lett 11(5):463–466. https://doi.org/10.1109/LSP.2004.824029
21. Shylaja C, Rai A, Mishra PK (2021) Modelling and simulation of 16-bit vedic multiplication using FPGA. J Phys Conf Ser 2007(1). https://doi.org/10.1088/1742-6596/2007/1/012003

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

Amrita Rai[1] · Ajay Roy[2] · Shamimul Qamar[3] · Abdulelah G. F. Saif[3] · Magdi Mohammad Hamoda[3] · Abdul Azeem[4] · Salman Arafath Mohammed[5]

Amrita Rai
amritaskrai@gmail.com

Ajay Roy
ajoy.22652@lpu.co.in

Abdulelah G. F. Saif
absaif@kku.edu.sa

Magdi Mohammad Hamoda
mahmedhamoda@kku.edu.sa

Abdul Azeem
abdul182564@st.jmi.ac.in

Salman Arafath Mohammed
salman@kku.edu.sa

1 Department of Electronics and Communications, GL Bajaj Institute of Technology and Management/ LLOYD Institute of Engineering and Technology, Greater Noida, UP, India

2 Lovely Professonal University, Jalandhar, India

3 Computer Science & Engineering, College Of Sciences & Arts, King Khalid University(Abha), Dharan, Al Janub Campus, Abha, Saudi Arabia

4 Electrical Engineering Department, JMI University/ Delhi Technological University, Delhi, India

5 Electrical Engineering Department (Computer Engineering Section), King Khalid University, Abha, Saudi Arabia

Physica Scripta

INCLUSIVE PUBLISHING
TRUSTED SCIENCE

PAPER

Novel hybrid-CMOS inverter utilizing phase transition material for enhancing digital logic performance at lower operating voltages

To cite this article: Ritika Sorot *et al* 2024 *Phys. Scr.* **99** 035024

View the article online for updates and enhancements.

You may also like

- Research on No-affected Photovoltaic Grid-Connected Inverter by the Inductance Value
Rongfei Zhang, Zichao Guan, Ying He et al.
- Single stage Boost Cascaded Multilevel Inverter Based on Switched Inductor Structure
P. Sriramalakshmi, Rahul Singh and Sneha
- Analysis of Total Harmonic Distortion and implementation of Inverter Fault Diagnosis using Artificial Neural Network
T G Manjunath, A C Vikramathithan and H Girish

This content was downloaded from IP address 14.139.251.98 on 21/02/2024 at 06:02

Physica Scripta

PAPER

Novel hybrid-CMOS inverter utilizing phase transition material for enhancing digital logic performance at lower operating voltages

Ritika Sorot[1], **Anubha Goel**[2] **and Sonam Rewari**[1]

[1] Electronics & Communication Engineering, Delhi Technological University, New Delhi, India
[2] Electronics & Communication Engineering, Maharaja Agrasen Institute of Technology, New Delhi, India

E-mail: anubhagoel15@gmail.com

RECEIVED
24 July 2023

REVISED
11 January 2024

ACCEPTED FOR PUBLICATION
7 February 2024

PUBLISHED
19 February 2024

Keywords: phase transition material, hybrid phase transition FET, insulating to metallic current density ($J_{C\text{-}IMT}$), metallic to insulating current density ($J_{C\text{-}MIT}$), insulating resistivity (ρ_{INS}), metallic resistivity (ρ_{METAL})

Abstract

This research paper introduces a novel design for a hybrid-CMOS inverter using vanadium dioxide (VO_2), a phase transition material. The proposed inverter exhibits a remarkably steep transition for falling logic at the output(1–0). By leveraging the insulating to metallic current density ($I_{C\text{-}IMT}$) of VO_2, the depth and gain of this transition can be finely tuned. Notably, a higher $I_{C\text{-}IMT}$ value yields a greater gain in the transition slope. In comparison to a traditional CMOS inverter, the designed inverter demonstrates several advantages. It achieves higher values of lower noise margin (NM_L) and significantly reduces static power dissipation by 97.7%. These promising outcomes present an exciting opportunity for designing inverters at lower drain voltages, especially in devices operating at lower technology nodes. Furthermore, the hybrid-CMOS inverter is designed to excel in the sub-threshold region of operation, resulting in elevated values of lower noise margin and reduced leakage current values.

1. Introduction

Phase transition materials are transition metal oxides like VO_2, chalcogenides, VO, TiO, doped chalcogenide, Cu-doped HfO_2, etc which have unique properties that can affect the behavior of transistors [1]. A phase transition material has the ability to exist in both a highly resistive insulating state and a very low resistive metallic state. The transition from the insulating state to the metallic state is initiated by the insulating-to-metallic critical current ($I_{C\text{-}IMT}$). Once the current begins to decrease and reaches the metallic-to-insulating critical current ($I_{C\text{-}MIT}$), the material switches back to the metallic state [2]. Considerable research is currently underway to investigate the practicality of phase transition material (PTM) in creating low-power circuits [3]. Notably, recent studies have revealed the potential of employing PTM in designing low-power circuits specifically tailored for medical devices [4]. A recent study has demonstrated the suitability of PTMs in the development of low-power dynamic circuits [5]. Moreover, it has been revealed that combining PTM with the negative capacitance effect improves the sub-threshold slope further and holds great promise for applications in digital devices and memories [6, 7].

By connecting a PTM to the source of field effect transistor(FET), hybrid phase transition FET (hyper-FET) with unconventional device characteristics can be created (figure 1). The bistable switching states of PTM make it a preferred option for application in neuromorphic computation [8]. Transfer characteristics and output characteristics of hyper-FET depend upon the parameters of PTM [9]. The integration of PTM into a transistor allows for faster switching speeds and lower power consumption that traditional metal oxide semiconductor field effect transistors (MOSFETs) may not possess [10]. Recent experimental research in steep switching devices has indicated the existence of hyper-FETs with sub-threshold slopes less than 10 mV/decade [11]. To take advantage of the steep sub-threshold slope of hyper-FET, it is required to find the optimized values of PTM parameters [12]. The proposed design of a hybrid-CMOS inverter using hyper-FET offers faster switching speed and better dc response than a conventional complementary metal oxide semiconductor(CMOS) inverter.

© 2024 IOP Publishing Ltd

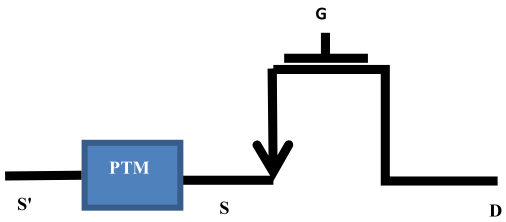


Figure 1. Hyper-FET symbol where S', G and D are source, gate and drain terminal of hyper-FET. S is the source of MOSFET.

There are several constraints associated with traditional CMOS inverter especially at lower technology nodes due to reduced operating voltages. Limitations encountered with conventional CMOS inverters are as given in the following points [2].

1. The sub-threshold slope of MOSFET is limited to 60 mV/decade. Since the sub-threshold slope determines the transistor's efficiency in turning off completely, a steeper slope is desirable to minimize leakage current. A shallow sub-threshold slope results in increased power consumption when the transistors are in an off state.
2. The sub-threshold slope affects the transition speed of the inverter, causing a slower response time and reducing the voltage levels at which the inverter can safely operate. This decreases the noise margin, making the circuit more susceptible to noise interference and affecting its overall performance.
3. A shallow slope leads to longer delay times, reducing the speed of the circuit. This limitation becomes more pronounced in complex digital circuits and can limit their maximum operating frequency.

In this paper, we propose a steep slope hybrid-CMOS inverter that utilizes a PTM. Leakage current, static power dissipation, and propagation delay are reduced to a great extent due to the steep transition and lower sub-threshold slope of hyper-FET.

In section (2.1), the existing modeling styles of phase transition material are discussed along with the physics of phase transition material. Section (2.2) discusses the deficiencies of the PTM-based hyper-FET inverter, which led to the design of the hybrid-CMOS inverter described in section (2.3). In the subsequent part, section (2.4) an in-depth analysis is undertaken to study the dc response of the hybrid CMOS inverter. This examination includes the evaluation of noise margin and a comparison with conventional CMOS inverters. Section (3) focuses on the comparison of static power dissipation of a hybrid-CMOS inverter with a conventional CMOS inverter while section (4) deals with the application of a hybrid-CMOS inverter for the sub-threshold region of operation. Finally, section (5) discusses the switching speed of a hybrid-CMOS inverter in comparison with a conventional CMOS inverter.

2. Design of hybrid-CMOS inverter

2.1. Modeling of phase transition material

Spice-compatible PTM models for simulating vanadium dioxide(VO_2) are currently unavailable. Ongoing experimental studies aim to derive analytical models based on the device's underlying physics.

Some authors associate the phase transition in VO_2 with the Mott-Hubbard scenario, while others link it to electron–phonon coupling, specifically a Peierls mechanism, based on changes in crystal symmetry. It has been suggested that VO_2 exhibits a combined nature, displaying characteristics of both Mott and Peierls transitions [13]. The transition can be triggered by an increase in free charge carrier density through photo generation, injection, or high-field generation during switching.

This initiation process, utilizing any of the aforementioned methods, can occur without the need for doping, pressure, or heating to a specific transition temperature. The metal-insulator transition (MIT) mechanism in VO_2 is dual, incorporating elements of both Mott and Peierls transitions. The initiating mechanism remains a correlation-driven electronic Mott transition. A field-induced rise in charge density serves to screen Coulomb interactions, ultimately eliminating the Mott-Hubbard energy gap at temperatures below T_t. This non-thermal, electronic switching mechanism in VO_2 should be taken into account when developing an analytical model [14].

It has been shown that based on Mott criterion for MIT, $a_H n_c^{1/3} \sim 0.25$, where a_H is the Bohr radius and n_c = critical charge density for MIT. Correlation length ξ is given by $2\hbar\nu_F/\Pi\Delta$, where, $\hbar$ is the Planck Constant, ν_F is electron Fermi velocity and Δ is the correlation energy gap. Bohr radius is, $a_H = \varepsilon\hbar^2/m^*e^2$where ε is material

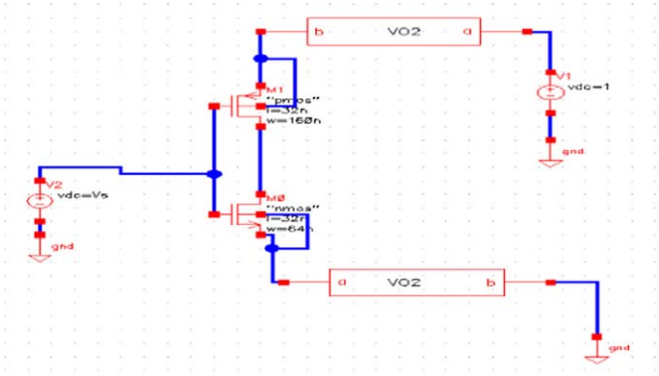


Figure 2. Schematic Of hyper-FET inverter.

Table 1. Parameters of PTM_P & PTM_N for hyper-FET inverter.

S.No	Parameters of PTM	Values of parameters
1.	$J_{C\text{-}IMT}$	520 A cm^{-2}
2.	$J_{C\text{-}MIT}$	8000 A cm^{-2}
3.	ρ_{INS}	149 Ωcm
4.	C	0.00149 Ωcm
5.	R_{INS}	22.6 MΩ
6.	R_{METAL}	226 Ω
7.	A_{PTM}	64*33 nm^2

dielectric constant, m^* is charge carrier effective mass and e is the charge of an electron. The relation is used to calculate the correlation length in vanadium dioxide which is found to be around 20A and 1–2A. MIT in VO_2 is two-stage (dual). The driving force of MIT in VO_2 is the electronic Mott-Hubbard transition [15]. The modulation of the electronic spectrum results then in distortion of the crystal structure accompanied by the formation of the gap and by the complete localization of electrons onto vanadium atoms with the localization radius. MOTT Transition is a trigger mechanism that then initiates the structural transformation. This literature related to vanadium dioxide can be used to make an analytical model [15].

As the analytical modeling is still in progress, the authors have chosen to use equations (1) and (2) to perform behavioral modeling of vanadium dioxide through verilog-A [15–17]. Here, R_{INS} is the resistance of PTM in the insulating state and R_{METAL} is the resistance of the PTM in the metallic state. $I_{C\text{-}IMT}$ denotes the insulating to metallic current density and $I_{C\text{-}MIT}$ denotes metallic to insulating current density. $V_{C\text{-}IMT}$ and $V_{C\text{-}MIT}$ denote the critical voltages required by PTM for the transition from an insulating state to a metallic state and vice versa.

$$V_{C-IMT} = I_{C-IMT}{}^{*}R_{INS} \tag{1}$$

$$V_{C-MIT} = I_{C-MIT}{}^{*}R_{METAL} \tag{2}$$

The verilog-A model is constructed utilizing parameters provided by different researchers [18–21] from their experiments on vanadium dioxide films. This verilog-A model of PTM can be integrated into cadence virtuoso schematic along with other circuit components to design logic circuits and conduct simulation studies.

2.2. Limitations of hyper-FET inverter

Hyper-FET is comprised of a MOSFET with its source connected to the PTM (figure 1). The Schematic of the hyper-FET inverter(figure 2) consists of both complementary N and P hyper-FETs that were connected to power supply rails via PTM [18]. The circuit of hyper-FET is designed in cadence virtuoso using FETs of 32 nm technology and a verilog-A model of vanadium dioxide for size 64 nm*33 nm. Table 1 lists the other parameters of PTM used for simulation [18]. The dc response curve of the hyper-FET inverter, as depicted in figure 3, is measured by performing dc analysis and simulation using the analog design environment (ADE) within the Cadence circuit simulation tools. In the beginning, when V_{in} is at a logic low level and both the PTMs are in insulating state, there is very little drain current flowing through the hyper-FETs. As the pull-up and pull-down resistances are both high, the output voltage (V_{out}) relative to V_{dd} is lower when the input voltage (V_{in}) is zero [18]. When the drain current flowing through the hyper-FETs and PTMs reaches the critical current ($I_{C\text{-}IMT}$),

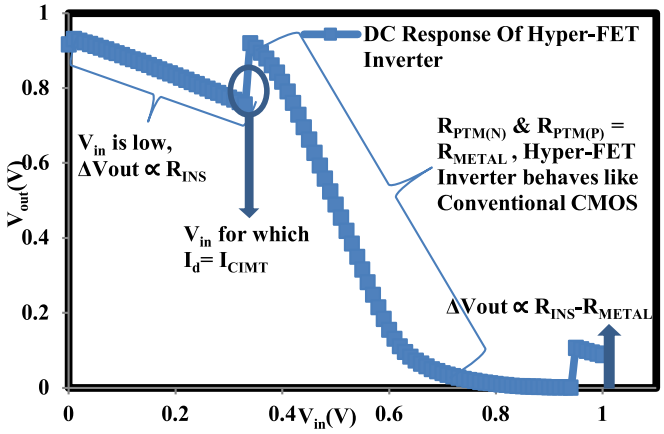


Figure 3. DC response illustrating the variation of output voltage (Vout) with input voltage(V_{in}) for hyper-FET inverter.

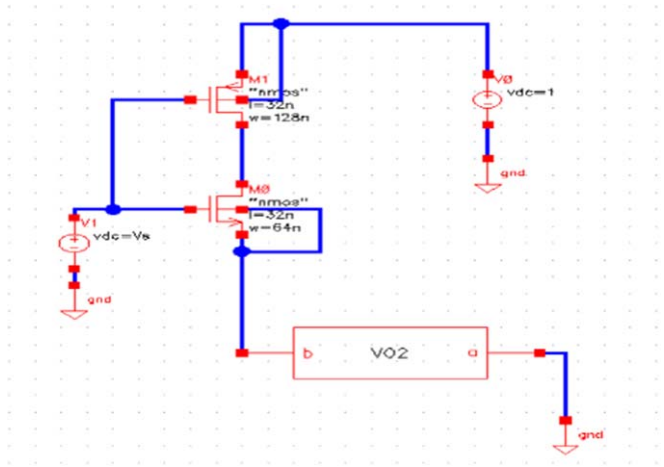


Figure 4. Schematic of hybrid-CMOS inverter.

the resistance of the PTM changes to R_{METAL}, causing an increase in current flow through the hyper-FET inverter and resulting in an output voltage drop. However, it is worth noting that the logic (1–0) transition of output is not steep.

Indeed, the response closely resembles that of a conventional CMOS inverter. As the input voltage approaches Vdd, the output voltage drops to zero but exhibits a slight rise. This phenomenon occurs because the PTM reverts into the metallic state when the current flowing through it falls below $I_{C\text{-}MIT}$. Notably, hysteresis is observed in the non-transition region of the dc response [18]. Comparatively, the hyper-FET inverter does not demonstrate an improvement in the transition slope when compared to the conventional CMOS inverter.

2.3. Circuit design of proposed hybrid-CMOS inverter

To minimize the relative values of the pull-up and pull-down resistances in the insulating state of the PTM and achieve a sharp falling output transition, it is essential to exclusively integrate the PTM with the source of the NMOS transistor. The PTM is not connected with the source of the PMOS transistor. As a result, the proposal suggests connecting hyper-N with the PMOS transistor to create a high-gain hybrid-CMOS inverter (figure 4). In this specific design, a verilog-A model of vanadium dioxide (VO_2) is utilized as the phase transition material. The parameters of PTM used for designing the hybrid-CMOS inverter have been documented in table 2, showing the alterations made from those originally provided in table 1. The parameters follow all the device design-related constraints [3, 9]. The circuit design in figure 2 is implemented and simulated in the analog design environment of the cadence tool.

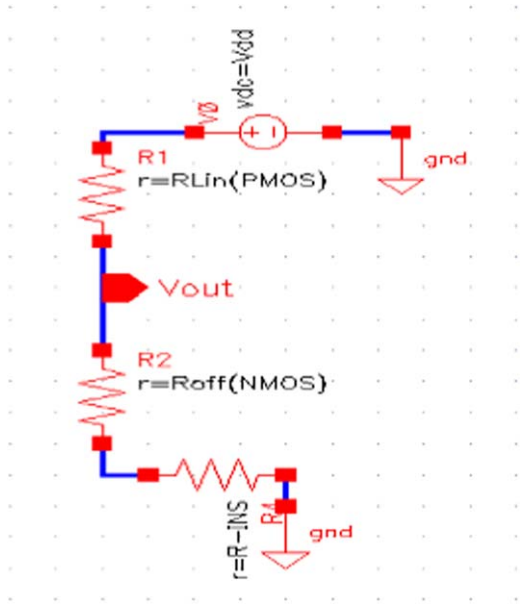


Figure 5. Equivalent circuit of Hybrid -CMOS inverter corresponding to region 1.

Table 2. Parameters of PTM for hybrid CMOS inverter.

S.no	Parameters of PTM	Values of parameters
1.	I_{C-IMT}	2.14E–09 A
2.	I_{C-MIT}	32.92E–09 A
3.	ρ_{INS}	6.86E + 12 Ω-cm
4.	P_{METAL}	6.86 E + 7 Ω-cm

2.4. Circuit design of proposed hybrid-CMOS inverter

To minimize the relative values of the pull-up and pull-down resistances in the insulating state of the PTM and achieve a sharp falling output transition, it is essential to exclusively integrate the PTM with the source of the NMOS transistor. The PTM is not connected with the source of the PMOS transistor. As a result, the proposal suggests connecting hyper-N with the PMOS transistor to create a high-gain hybrid-CMOS inverter (figure 4). In this specific design, a verilog-A model of vanadium dioxide (VO_2) is utilized as the phase transition material. The parameters of PTM used for designing the hybrid-CMOS inverter have been documented in table 2, showing the alterations made from those originally provided in table 1. The parameters follow all the device design-related constraints [1].

2.5. Analysis of DC response of proposed hybrid-CMOS inverter

VO_2 remains insulating unless triggered by a critical current. We denote the voltage and current required for VO_2 to transition from the insulating state to the metallic state as $V_{C\text{-}IMT}$ and $I_{C\text{-}IMT}$, respectively. Conversely, $V_{C\text{-}MIT}$ and $I_{C\text{-}MIT}$ represent the voltage and current below which VO_2 transitions back from the metallic state to the insulating state. Table 2 in the paper provides the relevant parameters of VO_2 used in the cadence virtuoso simulation. During the design process, it is ensured that $V_{C\text{-}IMT}$ is greater than $V_{th,n}$, while also adhering to all constraints associated with device modeling. These measures are taken to achieve a steep decline in slope.

2.5.1. Analysis of DC response with increasing input voltage (logic 1–0)

Figure 6 illustrates the dc response curve of a hybrid-CMOS inverter for falling values of output (Logic 1–0). The following section elucidates the different regions observed in the DC response curves.

Region 1: When Input Voltage, V_{in} is between 0 V and V_{c1}(critical voltage at which the output voltage begins to drop($0 < V_{in} < V_{C1}$). This case corresponds to low values of V_{in}, such that $V_{VO2} < V_{C\text{-}IMT}$, $V_{in} < V_{th,n}$ & $I_{VO2} < I_{C\text{-}IMT}$. Figure 5 depicts the equivalent circuit for this region. VO_2 is in an insulating state represented by R_{INS} which is approximately 22.6 M-Ω. The drain current flowing through NMOS is not sufficient to trigger it into a metallic state. Therefore, PTM stays in a high resistive state. PMOS is in the linear region which is represented by $R_{Lin(PMOS)}$ [22, 23]. It is to be noted that although PMOS is in the linear region, still the drain current flowing

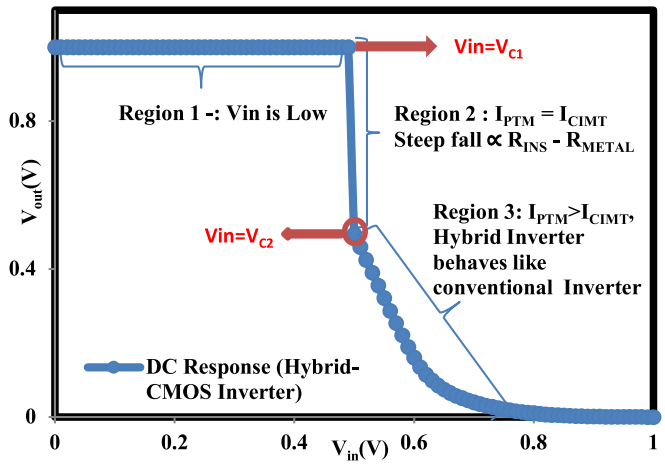


Figure 6. DC response depicting the variation of output voltage(V_{out}) with input voltage (V_{in})for hybrid-CMOS inverter.

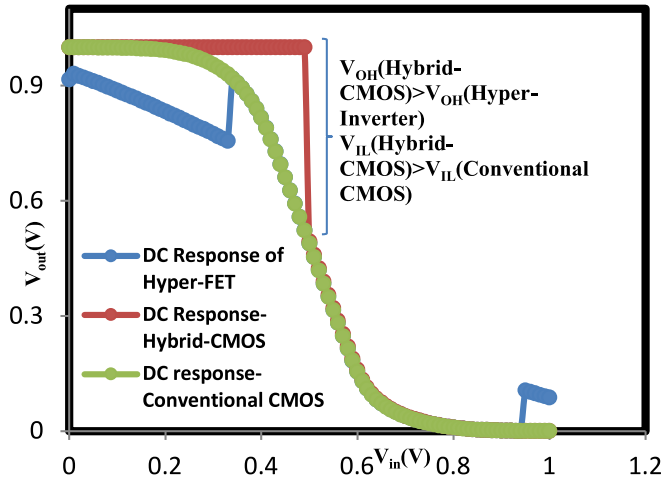


Figure 7. Comparison of DC response curves for Hyper-FET, hybrid-CMOS, and conventional CMOS inverter.

through it is negligible because NMOS in combination with PTM is off. Effective off-state resistance of hyper-N is equal to the addition of resistance of NMOS in off-state, $R_{off,n}$ [22, 23] and $R_{INS,}$ i.e. ($R_{off,n} + R_{INS}$) which is very much greater than $R_{Lin,p}$(resistance of PMOS in the linear region). Therefore, the drain current flowing through both the MOSFETs ($I_{d,n}$, $I_{d,p}$)is nearly zero Amperes i.e. $I_{dn} = I_{dp} \sim 0$A. Hence, the Voltage drop across $R_{Lin,p}$ is also approximately 0 V. This implies the output voltage of the inverter, $V_{out} \sim V_{dd}$. Figure 7 reveals that $V_{out(Hybrid)} > V_{out(CMOS)} > V_{out(Hyper)}$ for V_{in}(0 to V_{IL}). Also, the voltage drop seen in the dc response curve of the hyper-FET inverter is not seen in the dc response curve of the hybrid-CMOS inverter since the pull-down resistance is significantly higher compared to the pull-up resistance.

Region 2 ($V_{in} = V_{C1}$): Here, the voltage across vanadium dioxide, V_{VO2} is equal to $V_{C\text{-}IMT}$ and V_{in} is greater than $V_{th,n}$ (threshold voltage of NMOS). At this point, VO_2 switches state from insulating(R_{INS}) to metallic(R_{METAL}) because it is triggered by current equal to critical current density ($I_{PTM} = I_{CIMT}$) .

Reason for steep fall: Effective resistance of hyper-N transistor just before switching is equal to ($R_{off,n}$+R_{INS}). (Despite the gate to source voltage of the NMOS transistor($V_{GS,n}$) being greater than $V_{th,n}$, the NMOS remains off due to the significant resistance (R_{INS}). This high resistance impedes the increase of $I_{d,n}$ (drain current of the NMOS) even after reaching the threshold voltage for the NMOS transistor).

If $I_{d,n}$ is equal to $I_{C\text{-}IMT}$, then R_{PTM} becomes equal to R_{METAL}. Voltage change across VO_2 at this switching point is equal to ($I_{C\text{-}IMT}{}^{*}R_{INS}$—$I_{C\text{-}IMT}{}^{*}R_{METAL}$).

This voltage change corresponding to the difference in resistance of insulating and metallic states is sudden and happens instantly when V_{PTM} reaches $V_{C\text{-}IMT}$ or I_{PTM} becomes equal to $I_{C\text{-}IMT}$. The change in voltage across VO_2 is reflected as a fall in the output voltage of the hybrid-CMOS inverter.

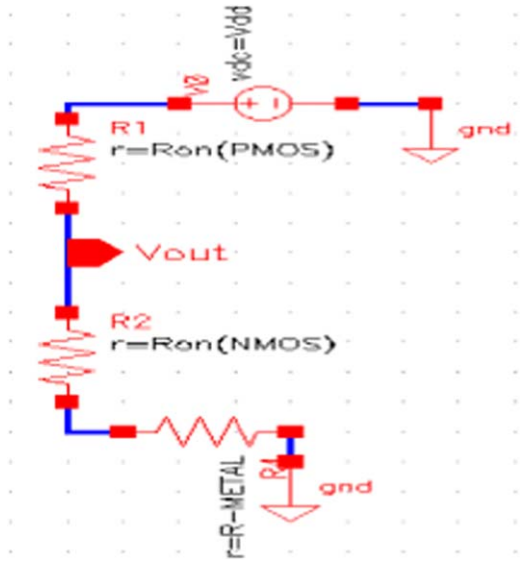


Figure 8. Equivalent circuit of Hybrid-CMOS inverter corresponding to region 3.

Calculation of V_{C1}: In the case of the hybrid-CMOS inverter, the output voltage does not exhibit a gradual decline but rather maintains the level of V_{OH} until this specific point. As a result, it is not possible to calculate V_{C1} using the conventional CMOS inverter's V_{IL} formula [24].

The node voltages of the NMOS indicate that it is operating in the saturation region, with $I_{C\text{-}IMT}$ representing the current flowing through VO_2 at that moment. To determine V_{C1}, the following equations can be utilized [24].

$$I_{d,sat,n} = I_{C-IMT} \tag{3}$$

$$I_{d,sat,n} = k/2(V_{GS,n} - V_{th,n})^2 \tag{4}$$

Here, $V_{in} = V_{C1}$, $V_G = V_{in}$, $V_S = V_{VO2}$

$$K/2(V_{in} - V_{VO2} - V_{th,n})^2 = I_{C-IMT} \tag{5}$$

$$V_{C1} = \sqrt{2I_{C-IMT}/K} + V_{th,n} + V_{VO2} \tag{6}$$

To achieve a significant decrease in the logic (1–0) delay of the output, it is crucial to carefully select the PTM parameters. The steep fall in the output voltage, resulting from the resistive state change of the PTM, should be at least half of the full voltage swing. If the voltage drop change is less than half of the full voltage swing, the logic delay reduction will be insignificant. For optimal results in terms of speed and Noise Margin, the PTM parameters should be chosen in a way that this steep fall is centered around the middle value of the input voltage range.

Region 3: When $V_{in} > V_{C1}$, NMOS & PMOS are in saturation & $R_{PTM} = R_{METAL}$. Equivalent circuit for this region is depicted in figure 8. R_{on} is the resistance of the MOS in saturation region.

Value of R_{METAL} is small and now the hybrid-CMOS inverter behaves like conventional CMOS inverter. Consequently, following the abrupt drop in the output voltage, the drain current adheres to the MOSFET drain current equation applicable to the saturation region. As a result, the output gradually diminishes towards zero.

Calculation of V_{C2}: The threshold at which the input voltage begins to exhibit CMOS inverter behavior varies depending on the value of $I_{C\text{-}IMT}$. The current flowing through VO_2 at this juncture can be determined by leveraging the fact that the voltage drop across VO_2 remains constant at the transition point.

$$\text{Therefore, } I_{C-IMT}{*}R_{INS} = I_C{*}R_{METAL}$$

$$I_C = (I_{C-IMT}{*}R_{INS})/R_{METAL} \tag{7}$$

Since NMOS is in saturation. The current I_c can be equated to the saturation current of NMOS corresponding to V_{in}=V_{C2} [24]

$$I_{d,sat,n} = I_C \tag{8}$$

$$I_{d,sat,n} = k/2(V_{GS,n} - V_{th,n})^2 \tag{9}$$

Here, $V_{in} = V_{C2}$, $V_G = V_{in}$, $V_S = V_{VO2}$

$$K/2(V_{in} - V_{VO2} - V_{th,n})^2 = I_{C-IMT} \tag{10}$$

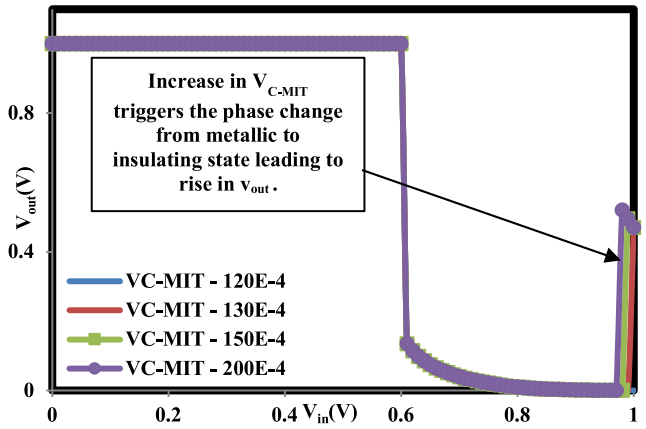


Figure 9. DC response curve of hybrid-CMOS depicting the relation between output voltage(V_{out}) and input voltage (V_{in}) for varying values of $I_{C\text{-}MIT.}$.

$$V_{C1} = \sqrt{2I_C/K} + V_{th,n} + V_{vo2} \tag{11}$$

By utilizing equations (6) and (11), it can be interpreted that the difference between V_{C2} & V_{C1} is proportional to the difference between $\sqrt{I_{C\text{-}IMT}}$ and $\sqrt{I_C}$.

$$V_{C2} - V_{C1} = \sqrt{\left(\frac{2I_C}{K}\right)} - \sqrt{(2I_{C-IMT}/K)} \tag{12}$$

In the scenario where the current flowing through PTM, immediately after being switched to the metallic state, is equal to $I_{C\text{-}IMT}$, V_{C1} and V_{C2} will be approximately equal. This condition leads to an approximate infinite gain of the transition slope.

2.5.2. Analysis of DC response with varied values of $I_{C\text{-}MIT}$/$V_{C\text{-}MIT}$ (logic 0–1)

When I_{PTM} decreases to a value lower than $I_{C\text{-}MIT}$, it switches back to the state of R_{INS}, resulting in a slight increase in the output voltage, as depicted in figure 9. It is crucial to maintain the value of $I_{C\text{-}MIT}$ in such a way that it is only reached when the input voltage surpasses the upper limit of the input voltage range. If the values of $I_{C\text{-}MIT}$ and $V_{C\text{-}MIT}$ are increased, the output voltage will rise slightly again after reaching the zero level, causing the output node to remain floating. This increase in V_{out} is equal to $I_{PTM}{}^{*}(R_{INS}\text{-}R_{METAL})$.

If the value of $I_{C\text{-}MIT}$ cannot be reduced due to device constraints, an alternative approach is to reduce the input voltage range. Reducing the input voltage range has minimal impact on the logic high and logic low levels of both the input and output voltage, as well as the noise margin.

2.5.3. DC Response for varying values of $I_{C\text{-}IMT}$

The effect of increasing $I_{C\text{-}IMT}$ values (for a fixed $I_{C\text{-}MIT}$ value) on the dc response curve of the hybrid-CMOS inverter is depicted in figure 10. The steep decline of the output voltage is directly proportional to $I_{C\text{-}IMT}$. Consequently, if the value of $I_{C\text{-}IMT}$ is very low, the change in V_{out} will be minimal. For extremely low $I_{C\text{-}IMT}$ values, the transition from logic high to logic zero resembles that of a conventional CMOS inverter.

At an $I_{C\text{-}IMT}$ value of 3.50E-9, the fall in output voltage corresponds to the difference between V_{OH} and V_{OL}. However, increasing $I_{C\text{-}IMT}$ further results in performance degradation of the hybrid-CMOS inverter, with V_{out} never reaching V_{OL}. Careful consideration must be given to choosing $I_{C\text{-}IMT}$ to optimize the steep fall and maximize the value of noise margins.

2.5.4. Analysis of DC response with decreasing input voltage(logic 1–0)

The dc response for falling values of V_{in} differs from that of rising input voltage, as shown in figure 11. Initially, the PTM remains in the insulating state. However, when a high input voltage is applied and the current surpasses $I_{C\text{-}IMT}$, the PTM transitions to the metallic state. This explains the high output voltage observed at the beginning when a high logic level is applied to the input.

As the input voltage starts to decrease, the current flowing through the PTM also decreases. The PTM remains in the metallic state until I_{PTM} reaches the value of $I_{C\text{-}MIT}$, which occurs when V_{PTM} approaches zero. To avoid generating oscillations in the output, $I_{C\text{-}MIT}$ is kept low and close to zero input voltage. This ensures that the transition to the insulating state occurs near zero input voltage. Up until that point, the curve of the dc response follows that of a conventional CMOS inverter.

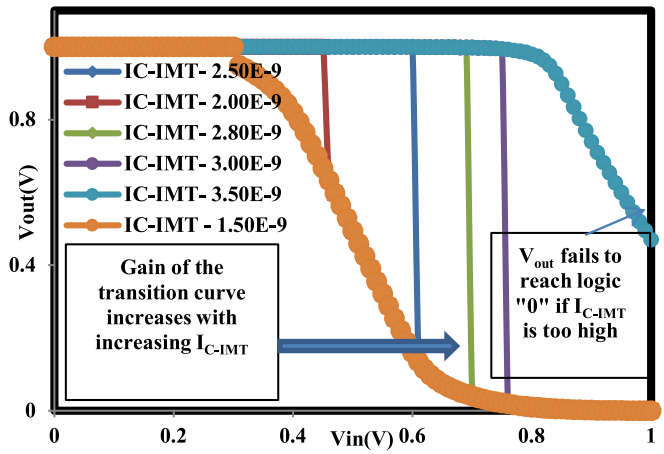


Figure 10. DC response of hybrid-CMOS inverter illustrating the relation between output voltage(V_{out}) and input voltage (V_{in}) across different values of $I_{C\text{-}IMT}$.

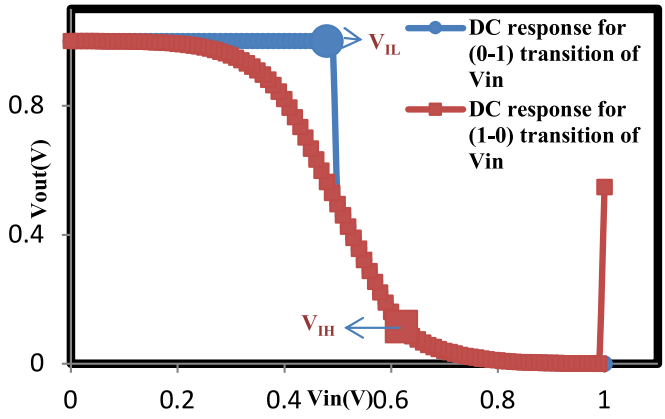


Figure 11. DC response of hybrid-CMOS inverter depicting the variations in output voltage(V_{out}) for (0–1) & (1–0) transition of V_{in}.

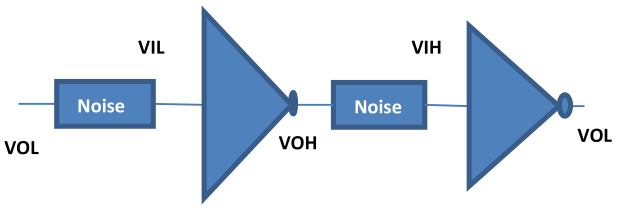


Figure 12. Series connected hybrid-CMOS inverters for calculating noise margin.

2.5.5. Noise margin for hybrid-CMOS inverter

Due to the dissimilarity between the DC response for falling and rising input values, the determination of V_{IL} and V_{IH} in the hybrid CMOS inverter differs from that of a conventional CMOS inverter. V_{OH} is the $V_{OUT,max}$ when the output level is logic '1'. According to the definitions for CMOS inverter [24], V_{OL} is the $V_{OUT,min}$ when the output level is logic '0'. V_{IL} is $V_{IN,max}$ which can be interpreted as logic '0'. V_{IH} is $V_{in,min}$ which can be interpreted as logic '1'.

Referring to figure 12 [24], in the case of the second inverter, when the input is logic '0', the output produces logic '1'. For the correct output of logic '0', V_{OL} needs to degrade to V_{IL}. By considering the dc response curve of the hybrid inverter for the logic (1–0) transition at the output, the correct logic will be produced at the output of the second inverter until V_{IL} is lower than V_{C1}.

$$\text{Noise margin(Low): } V_{IL} - V_{OL} = V_{C1} - V_{OL}$$

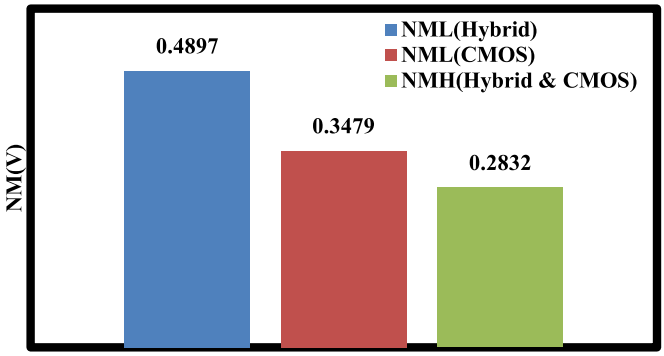


Figure 13. Noise Margin of hybrid-CMOS inverter & conventional CMOS inverter.

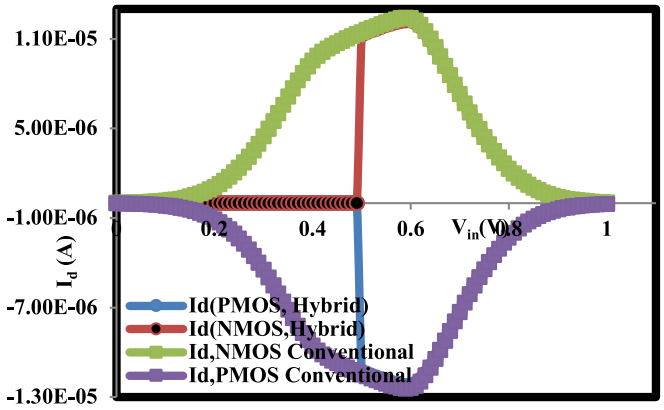


Figure 14. Drain current comparison of hybrid-CMOS inverter & conventional CMOS inverter for DC analysis.

$$NML = V_{IL} - V_{OL} \sim 489.75 \text{ mV}$$

The Noise margin low is higher for the hybrid-CMOS inverter compared to the conventional CMOS inverter due to the higher value of V_{C1}, as illustrated in figure 13. In the case of the third inverter, the input V_{OH} results in an output V_{OL}. However, this V_{OL} will persist until V_{OH} degrades to the level of V_{IH}. As the rising output of the hybrid-CMOS inverter resembles that of the conventional inverter, the value of V_{IH} is similar for both inverter designs.

NM_H is equivalent to V_{OH}-V_{IH} [24], which remains the same for both the conventional CMOS inverter and the hybrid-CMOS inverter, as depicted in figure 13. The values of NM_L can be tuned by adjusting the parameters of VO_2. Decreasing the value of the critical current, $I_{C\text{-}IMT}$, while keeping other parameters constant, will cause the noise margins to approach those of the conventional CMOS inverter. Conversely, increasing the value of $I_{C\text{-}IMT}$ leads to an increase in noise margin low (NM_L). Therefore, the steeper the slope, the higher the NM_L.

2.5.6. Static power dissipation of hybrid CMOS inverter

The current flowing through the hybrid-CMOS inverter remains minimal until reaching the point of $V_{C\text{-}IMT}$. Despite being in saturation, the NMOS transistor does not conduct due to the presence of a high-value insulating resistance (R_{INS}) connected to it. As depicted in figure 14, the drain current is approximately zero for lower input voltages. This characteristic leads to nearly zero power dissipation, as illustrated in figure 15.

It is essential to consider static power dissipation to optimize power dissipation in CMOS circuits, especially in low-power applications. Static power dissipation is calculated as the product of I_{static} & V_{DD} (Supply Voltage) [22]. I_{static} is equal to the sum of leakage currents of NMOS and PMOS [22, 23]. Static power dissipation mainly comes from leakage currents of the MOSFETs which occur even when there is no switching activity in the circuit. Table 3 presents the modified parameters of PTM utilized in the design of the hybrid CMOS inverter, deviating from the values initially provided in table 1. A comparison between the hybrid-CMOS and conventional CMOS reveals that the addition of PTM to the CMOS circuit reduces the static power dissipation by 97.7% as tabulated in table 4. Hence, this concept can be employed in the design of digital circuits for lower drain voltages and sub-threshold circuits, as it provides improved noise margin.

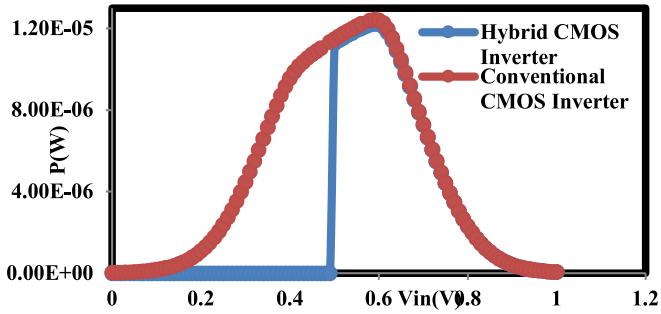


Figure 15. Power dissipation of hybrid-CMOS inverter & conventional CMOS inverter for DC analysis.

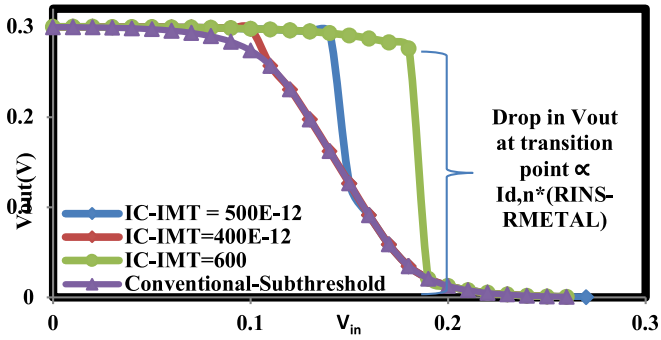


Figure 16. Comparison between sub-threshold DC Response of hybrid-CMOS inverter & conventional CMOS inverter.

Table 3. Parameters of PTM for sub-threshold operation.

S. no	Parameters of PTM	Values of parameters
1.	I_{C-IMT}	500E–12 A
2.	I_{C-MIT}	7.69E–09 A
3.	ρ_{INS}	6.86E + 12 Ω-cm
4.	P_{METAL}	6.86 E + 7 Ω-cm

3. Design of sub-threshold hybrid CMOS inverter

The sub-threshold operation involves smaller voltage swings compared to strong inversion operation. The reduced voltage range limits the seperation between logic high and logic low levels, which decreases the noise margin. As a result, the circuit becomes more susceptible to noise interference and may experience signal degradation or erroneous behavior. Sub-threshold inverters based on CMOS logic offer benefits such as reduced power consumption and potential for low voltage operation, but they are associated with a trade-off in terms of low noise margin.

The proposed hybrid CMOS inverter was designed for sub-threshold operation. Table 3 presents the modified parameters of PTM utilized in the design of the hybrid CMOS inverter, deviating from the values initially provided in table 1. The obtained results demonstrate a steep transition (figure 16) for falling logic at the output (1–0), resulting in enhanced noise margin (figure 17). Also the figures 17 and 18 depict that static current and static power dissipation decrease when critical current is set at higher value. Reduced drain current (figure 19) and power dissipation (figure 18) when compared to conventional CMOS inverters.

4. Switching speed of hybrid-CMOS inverter

The switching speed of the hybrid-CMOS inverter can be approximated by determining the propagation delay through transient simulation of the circuit in Cadence, using the same methodology as employed for calculating the switching speed of a conventional CMOS inverter [23–25]. Propagation delay, $\tau_P = (\tau_{PLH} + \tau_{PHL})/2$ where

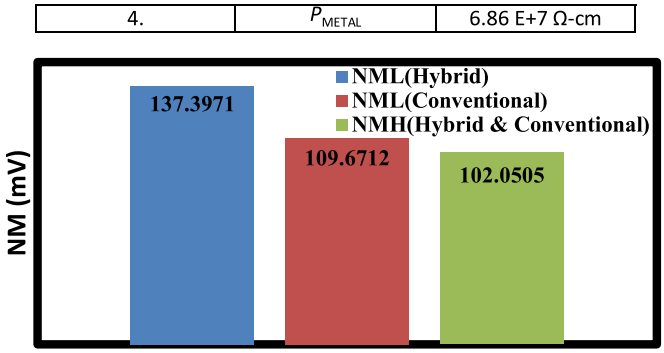


Figure 17. Noise Margin of Hybrid-CMOS & Conventional CMOS for sub-threshold region of operation.

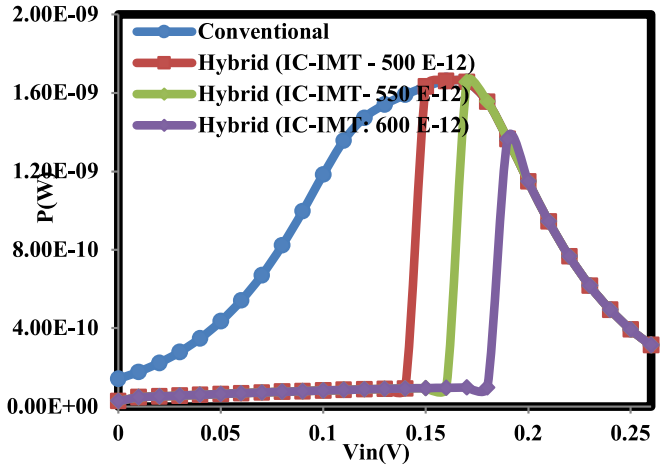


Figure 18. $I_{C\text{-}IMT}$ modulated static power comparison of Hybrid-CMOS Inverter & conventional CMOS Inverter for subthreshold region of operation.

Table 4. Comparison of conventional CMOS & hybrid-CMOS.

S. no	Parameter	Conventional CMOS (32 nm)	Hybrid-CMOS (32 nm)	Observation
1	Static Power	53.39 nW	1.192 nW	Static Power Of Hybrid-CMOS reduced by 97.7%.
2.	Noise Margin (Low)	0.3479 V	0.4897 V	NM_L for Hybrid-CMOS increases by 40.75%
3	Noise Margin (High)	0.2832	0.2832	Remains the same for both Inverters.

τ_{PHL} is the delay for the output to go from logic '1' to logic '0' and τ_{PLH} is the delay for the output to go from logic '0' to logic '1'.

Observing figure 20, it is noteworthy that when a waveform with a period of 200ps is applied as input, the propagation delay of the hybrid-CMOS inverter is 7.37ps, surpassing that of the conventional CMOS inverter with a propagation delay of 2.87ps. Although the value of τ_{PLH} is slightly lower for the hybrid-CMOS inverter, this increase can be attributed to the elevated τ_{PHL} value, as illustrated in figure 20(b), where the output voltage remains stable in the hybrid-CMOS inverter due to the insulating stage of PTM. In figure 21, a comparison of the propagation delay between the hybrid-CMOS inverter and the conventional CMOS inverter is presented when applying an input waveform with a period of 100ns. The results show a similar propagation delay in this case. Furthermore, in figure 22, when an input signal with a period of 50ps is applied, the hybrid-CMOS inverter behaves similarly to the conventional CMOS inverter, with a delay of approximately 13.34ps in both cases. Hence, it can be concluded that for optimized values of PTM parameters, the hybrid-CMOS inverter yields significant power savings across a broad range of input signal frequencies without substantial sacrifice in switching speed.

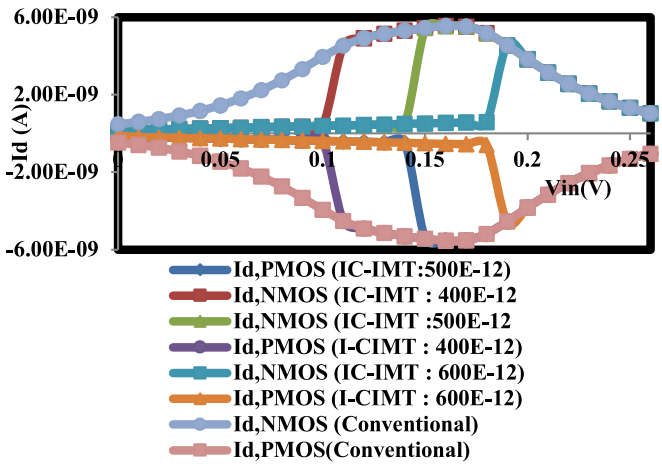


Figure 19. $I_{C\text{-}IMT}$ modulated drain current comparison of Hybrid-CMOS & Conventional CMOS Inverter for DC analysis.

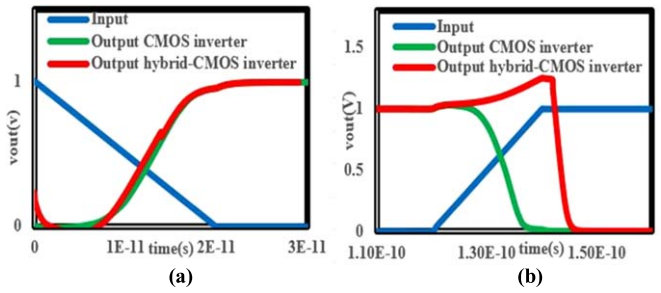


Figure 20. Comparison of conventional CMOS inverter with hybrid-CMOS inverter using an input waveform with a period of 200ps, considering (a) the delay of output waveform transitioning from logic '0' to logic '1' ($\tau_{PLH,}$) (b) the delay of ouput waveform transitioning from logic'1' to logic'0' ($\tau_{PHL,}$).

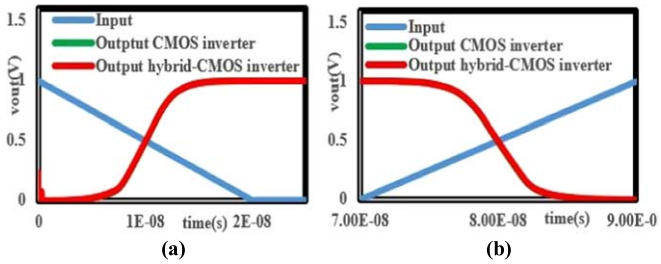


Figure 21. Comparison of conventional CMOS inverter with hybrid-CMOS inverter using an input waveform with a period of 100ns, considering (a) the delay of output waveform transitioning from logic '0' to logic '1' ($\tau_{PLH,}$) (b) the delay of ouput waveform transitioning from logic'1' to logic'0' ($\tau_{PHL,}$).

The switching speed is not affected by the high value of insulating resistivity. The resistivity of the phase transition material remains high only during its insulating state. Upon being triggered by the critical current into the metallic state, the resistivity undergoes a substantial decrease. Consequently, the switching speed is determined by the duration it takes for the phase transition material to shift from its insulating state to the metallic state, commonly referred to as the transition time of the phase transition material. Additionally, the PTM parameters such as $I_{C\text{-}IMT}$, $I_{C\text{-}MIT}$, R_{INS}, and R_{METAL} will influence the values of switching speed of hybrid-CMOS inverter. Optimizing these parameters is anticipated to decrease the propagation delay for some frequency range of the input signals.

Therefore, a detailed transient analysis is needed to study the impact of these parameters on the switching speed of hybrid-CMOS inverter.

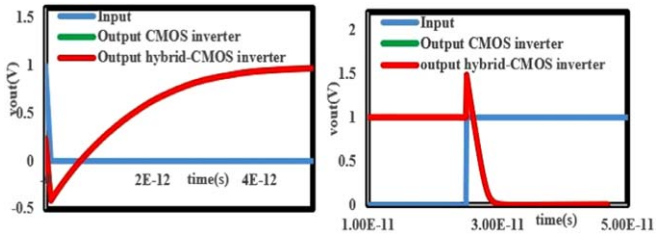


Figure 22. Comparison of conventional CMOS inverter with hybrid-CMOS inverter using an input waveform with a period of 100ns, considering (a) the delay of output waveform transitioning from logic '0' to logic '1' ($\tau_{PLH,}$) (b) the delay of ouput waveform transitioning from logic'1' to logic'0' ($\tau_{PHL,}$).

5. Conclusion

The incorporation of phase transition material in the proposed hybrid-CMOS inverter design enhances digital logic performance by amplifying the gain of the dc response curve during falling logic (1–0) at the output. This increased gain reduces the lower noise margin by 40.75% & static power dissipation by 97.7%. However, the addition of phase transition material to the source of FET introduces an area penalty when compared with the traditional FET. A trade-off will have to be made between power saving and area for the utilization of a hybrid-CMOS inverter.

Additionally, the design was investigated for sub-threshold operation at lower drain voltages (0.3 V for 32 nm technology), demonstrating favorable performance in the dc analysis. Consequently, this inverter proves well-suited for utilization at lower operating voltages, particularly in lower technology nodes.

Data availability statement

The data cannot be made publicly available upon publication because they are not available in a format that is sufficiently accessible or reusable by other researchers. The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Anubha Goel https://orcid.org/0000-0002-0400-5664
Sonam Rewari https://orcid.org/0000-0001-8946-9849

References

[1] Morin F J 1959 Oxides which show a metal-to-insulator transition at the neel temperature *Phys. Rev. Lett.* **3** 34
[2] Ko E, Shin J and Shin C 2018 Steep switching devices for low power applications: Negative differential capacitance/resistance field effect transistors *Nano Convergence* **5** 1–9
[3] Jung H, Chang J, Yoo C, Oh J, Choi S, Song J and Jeon J 2022 Hyper-FET's phase-transition-materials design guidelines for ultra-low power applications at 3 nm technology node *Nanomaterials* **12** 4096
[4] Mitra J and Mitra J 2022 Exploring the potential of metal oxides for biomedical applications *Metal Oxides for Biomedical and Biosensor Applications* (Elsevier) pp 183–203
[5] Islam M M, Rivero M H, Rose G and Aziz A 2022 Low-power dynamic circuit design with steep-switching hybrid phase transition FETs (hyper-FETs). *IEEE Trans. Electron Devices* **70** 819–25
[6] Yadav S, Upadhyay P, Awadhiya B and Kondekar P N 2021 Ferroelectric negative-capacitance-assisted phase-transition field-effect transistor *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **69** 863–9
[7] Yadav S, Upadhyay P, Awadhiya B and Kondekar P N 2020 Design and analysis of improved phase-transition FinFET utilizing negative capacitance *IEEE Trans. Electron Devices* **68** 853–9
[8] Lu H, Clark S, Guo Y and Robertson J 2021 The metal–insulator phase change in vanadium dioxide and its applications *J. Appl. Phys.* **129** 24
[9] Sorot R, Goel A and Rewari S 2023 Phase transition material modulated hyper FET for digital applications In*2023 IEEE Devices for Integrated Circuit (DevIC)* 261–5(Piscataway, NJ: IEEE)
[10] Núñez J and Avedillo M J 2018 Power and speed evaluation of hyper-FET circuits *IEEE Access* **7** 6724–32
[11] Jimenez M, Nunez J and Avedillo M J 2020 Hybrid-phase-transition FET devices for logic computation *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits* **6** 1–8
[12] Jiménez M, Núñez J and Avedillo M J 2020 An approach to the device-circuit co-design of hyperFET circuits In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)* (IEEE) 1–5
[13] Stefanovich G, Pergament A and Stefanovich D 2000 Electrical switching and Mott transition in VO2 *J. Phys. Condens. Matter* **12** 8837

[14] Pergament A, Crunteanu A, Beaumont A, Stefanovich G and Velichko A 2016 Vanadium dioxide: metal-insulator transition, electrical switching and oscillations. A review of state of the art and recent progress arXiv:1601.06246
[15] Shukla N, Thathachary A V, Agrawal A, Paik H, Aziz A, Schlom D G and Datta S 2015 A steep-slope transistor based on abrupt electronic phase transition *Nat. Commun.* **6** 7812
[16] Tsai W Y, Li X, Jerry M, Xie B, Shukla N, Liu H and Narayanan V (2016). Enabling new computation paradigms with HyperFET-an emerging device *IEEE Transactions on Multi-Scale Computing Systems* **2** 30–48
[17] Maffezzoni P, Daniel L, Shukla N, Datta S and Raychowdhury A 2015 Modeling and simulation of vanadium dioxide relaxation oscillators *IEEE Trans. Circuits Syst. I* **62** 2207–15
[18] Aziz A, Shukla N, Datta S and Gupta S K 2017 Steep switching hybrid phase transition FETs (hyper-FET) for low power applications: A device-circuit co-design perspective—Part II *IEEE Trans. Electron Devices* **64** 1358–65
[19] Hua Q, Gao G, Jiang C, Yu J, Sun J, Zhang T and Wu H 2020). Atomic threshold-switching enabled MoS2 transistors towards ultralow-power electronics *Nature Communications* **11** 6207
[20] Avedillo M J and Núñez J 2017 Insights into the operation of hyper-FET-based circuits *IEEE Trans. Electron Devices* **64** 3912–8
[21] Shin J, Ko E and Shin C 2017 Analysis on the operation of negative differential resistance FinFET with Pb (Zr 0.52 Ti 0.48) O 3 threshold selector *IEEE Trans. Electron Devices* **65** 19–22
[22] Yeap G K 2012 *Practical Low Power Digital VLSI Design* (Springer Science & Business Media)
[23] Weste N H and Harris D 2015 *CMOS VLSI Design: A Circuits And Systems Perspective* (Pearson Education India)
[24] Kang S-M 1945 *CMOS Digital Integrated Circuits : Analysis and Design* (McGraw-Hill) 1996
[25] Berglund C N and Guggenheim H J 1969 Electronic properties of V O 2 near the semiconductor-metal transition *Phys. Rev.* **185** 1022

Biomass Conversion and Biorefinery
https://doi.org/10.1007/s13399-024-05394-8

ORIGINAL ARTICLE

Optimization of acidic pre-treatment conditions using response surface methodology for ethanol production from *Pistia stratiotes* using *Saccharomyces cerevisiae* and *Pichia stipitis*

Sweeti Mann[1] · Jai Gopal Sharma[1] · Rashmi Kataria[1,2]

Received: 18 October 2023 / Revised: 19 January 2024 / Accepted: 31 January 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
Bioethanol is a renewable chemical that can replace fossil fuels. *Pistia stratiotes*, an aquatic weed, was used as the substrate for the synthesis of ethanol in the current investigation. The presence of high amount of cellulose (25.90%) and hemicellulose content (18.44%) makes it suitable to produce bioethanol. Optimization of acidic pre-treatment conditions was performed using response surface methodology where independent variables acid concentration (0.15–3.1% H_2SO_4), time (12.96–97.0 min) and temperature (93.18–126.82 °C) were taken. The responses were recorded as sugar and cellulose concentrations. Subsequently, optimized condition was subjected to fermentation for the production of ethanol using *Saccharomyces cerevisiae* and *Pichia stipitis* strains. The maximum amount of sugar was produced under optimal conditions (2.5% acid concentration) at 120 °C for 15 min. Initial sugar was taken as 10 g/L in the hydrolysate, which is consumed by the strains with maximum sugar conversion rate (0.24 g/L/h). Maximum sugar consumption was observed by *S. cerevisiae* and *P. stipitis* and was 85.9% and 87.9%, respectively. Thus, we present that the optimized acidic pre-treatment conditions for maximum yield of ethanol from *Pistia stratiotes* using *S. cerevisiae* and *P. stipitis* were observed to be 0.37 and 0.39 respectively. Hence, this study showed an effective conversion of plant biomass into ethanol.

Keywords Bioethanol · Fermentation · Independent variables · Pre-treatment · Optimization · Sugar

1 Introduction

In the present scenario, energy demands are fulfilled by fossil fuels, which are rapidly depleted with the release of an excess number of greenhouse gases like CO_2 and CO during combustion. To overcome the problem for future perspective, searching for fossil fuel alternatives, such as biobutanol, bioethanol and biodiesel, that work in a sustainable manner for a clean environment will fulfil the energy demand [1]. Renewable and sustainable energy resources gained focus as an alternative to fossil fuels due to their lesser impact on the environment. Biomass from renewable feedstock attains interest in the research and development field due to its composition that contains a high amount of carbohydrates, proteins and lipids and cost-effectiveness [2]. Ecofriendly substrate obtained from plants is considered a renewable energy source with carbon-free emission [3]. At the industrial level, bioethanol has been produced by corn, wheat, sugar beet and sugarcane which are first-generation biomass. However, these feedstocks may have an unfortunate effect on the farmland, causing conflict in food and fuel crop generation. Lignocellulosic biomass falls under the second-generation feedstock, such as certain invasive plants, weeds and other lignocellulosic biomass from terrestrial and aquatic ecosystems. This material could be used as a carbon source in an economic way for energy production. USA, Russia, China, Germany, Brazil and India have taken action to generate bioenergy products from agriculture and industrial waste [4]. The Indian National Policy on Biofuels

✉ Rashmi Kataria
rashmikataria@gmail.com; rashmi.kataria@vit.ac.in

Sweeti Mann
sweeti.dhankhar.1988@gmail.com

Jai Gopal Sharma
sharmajaigopal@gmail.com

[1] Department of Biotechnology, Delhi Technological University, Shahbad Daulatpur, Bawana Road, Delhi, India

[2] School of Biosciences and Technology, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India

Published online: 16 February 2024

in 2018 provides a suggestive goal of attaining 20% ethanol blending with petrol by 2030 [5].

Bioethanol production is challenging due to high enzyme cost and technological barriers, such as biomass recalcitrance, diligent pre-treatment and inhibitory product formed during pre-treatment, which decreases the fermentation efficiency. The polysaccharides, cellulose and hemicellulose, present in lignocellulosic biomass, constitute 60% of its total mass and are reluctant to break it into simple sugar naturally [6]. The complex structure of plant cell walls prohibits the degradation of plant biomass into simple sugar. Different steps such as retreatment, hydrolysis and fermentation are required to convert the plant biomass into valuable products [7].

In recent decades, a lot of research is conducted on the optimization of processes for maximizing the production of valuable products from lignocellulosic biomass. Animal dung, agricultural waste and organic waste are examples of lignocellulosic biomass, which is pre-treated before being transformed into biogas. Due to the complex material's pre-treatment degradation into simple sugars, the results were incredibly effective [8]. Various pre-treatment methods are used for aquatic weeds, but 1% NaOH (alkali) pre-treatment is used for *Pistia stratiotes*, which increases biogas production [9]. Effective pre-treatment methods are required to increase the accessibility of substrate to the enzymes by changing the complex lignocellulosic structure into a simpler form [10]. The development of systematic pre-treatment methods resolves the initial obstacle in cellulosic ethanol production [11]. Many pre-treatment methods, viz. physical, mechanical and chemical (acidic and alkaline etc.) pre-treatment, have been applied to different types of lignocellulosic materials. The main motive of these pre-treatment methods is to solubilize the cell wall components (cellulose, hemicellulose and lignin). Acidic pre-treatment is much more effective compared to other pre-treatment methods as it does not require hydrolysis of the sample to solubilization of sugar reducing the energy consumption for the process [12]. H_2SO_4 is commonly used in pre-treatment due to its effectiveness in the removal of hemicellulose and lignin [13]. During acidic pre-treatment, lignocellulosic biomass is heated by an autoclave with a range of 100–120 °C, and during this procedure, some toxic compounds were generated, which can be minimized by adding the different chemicals [14].

Aquatic weeds present in freshwater bodies cause complications in irrigation and aquaculture projects. These weeds grow rapidly in the presence of proper nutrients and restrict the penetration of light on the lower surface of water bodies, which impairs the growth of aquatic biota. So it has become a topic of research to use it as a renewable carbon source for the production of valuable products in a sustainable manner [15]. The main advantages of these aquatic weeds are (1) they are an economical carbon source, (2) they have a higher CO_2 diminution impact as compared to the terrestrial plants and (3) they possess lower lignin content as compared to the terrestrial plants [16]. Extensive studies are conducted for the production of bioethanol from aquatic weeds due to these advantages.

Pre-treatment optimization gives a good direction in the field of bioethanol production. Besides pre-treatment, other barriers in biofuel processing are the commercialization of lignocellulosic biorefinery, feedstock organization and ample water consumption for the cultivation of weeds [17]. The aquatic weed *Pistia stratiotes* is composed of high carbohydrate content and can grow rapidly in the water. Therefore, it was selected as a substrate for ethanol production in our study. It was established in the earlier research on *Pistia stratiotes* that this weed's dry biomass produces the best results following pre-treatment. Following acid treatment, alkaline treatment was administered [18]. In our study, we have performed the optimization of acidic pre-treatment on this weed to get the maximum sugar, which was not done previously by any researcher. The parameters used for acidic pre-treatment can be optimized with the help of response surface methodology (RSM) design by Minitab software. The main objective of this model is to find out the main response in a particular area of interest, optimize the response, achieve the specific condition using a minimum number of experiments and observe the interaction between the parameters [19]. Combination of mathematical and statistical methods in RSM is used for making the model formation by which optimization of factors is done. The RSM method takes quantitative data from different experiments to estimate the regression model for optimization of responses (dependent variable) affected by process variables (independent variables) [20]. Central composite design (CCD) is a fractional factorial design consisting of 2^n factorial runs with $2n$ axial runs and centre runs that decide the experimental error. Independent variables determine the number of runs in the model. If the number of variables increases, the number of runs also increases in the replicates of the model. The RSM method for optimization consists of three steps. First step is designing of experiment statistically, the second step is to calculate the coefficients in the model and the last is to determine the response and examine the competency of the model within the design of the experiment [21].

The sugar obtained from the optimized condition was used for ethanol production. *Saccharomyces cerevisiae* and *Pichia stipitis* both strains were used for production. *Pichia stipitis* can use both C5 (e.g. xylose) and C6 (e.g. glucose) carbon sources present in the acid hydrolysate. This research makes a genuine contribution to the advancement of the technology used in the production of ethanol and may help in the development of an economical and sustainable approach for the production of ethanol by acidic pre-treatment.

2 Materials and methods

2.1 Collection and preparation of biomass

Pistia stratiotes was collected from a pond located at Hauz Rani City Forest, Delhi, India (28.5159° N, 77.2111° E). Average plant size was 7.57 cm, and after the collection of a sample, it is washed with tap water 4 times and then air dried in shadow at room temperature for 80 h. Air dried sample was put in the oven for 3 h at 40 °C and further ground using an electric grinder. Grounded samples were sieved with particle size 425 µm and stored in an air-tight jar for further use at room temperature [22].

2.2 Reagents

All the reagents used in this study were of analytical gradient grade. Sulphuric acid used for the acidic pre-treatment was purchased from Fisher Scientific, India. For the ethanol analysis, HPLC grade ethanol, purchased from Sigma-Aldrich, USA, was used in GC analysis.

2.3 Composition analysis of the biomass

Composition of biomass is essential to determine its capability for bioethanol production. Physicochemical analysis was performed on the raw sample through proximate analysis. The characterization of the sample was done by Fourier-transform infrared (FTIR) spectroscopy, TGA (thermo-galvanometric analysis) and scanning electron microscopy (SEM). The FTIR analysis and TGA were done at the Delhi Technological University, Delhi (India), while SEM was carried out at AIIMS, Delhi (India). The details for these analyses are mentioned below.

2.3.1 Determination of ash, total solids and moisture content

All of these compositions were done by the NREL protocols. Ash content of oven dried sample was estimated by the muffle furnace. In a pre-heated crucible, 1 g sample was burned at 575 °C for 6 h. Then, the crucible was put into a desiccator until it cooled down, and weight was recorded [23, 24]. The moisture content and total solid content were estimated by oven-dry procedure. The crucible was pre-heated at 105 °C for 4 h, and its weight was recorded. One gramme sample was put in the crucible at 105 °C for 4 h, then the crucible was transferred to the desiccator for cooling and the final weight was recorded [24].

$$\text{Total solids } (\%) = \frac{(\text{weight}_{\text{dry pan+dry sample}}) - \text{Weight dry pan}}{\text{Weight sample as received}} \times 100$$

$$\text{Total moisture}(\%) = 100 - \text{total solids}$$

2.3.2 Determination of cellulose content

Cellulose content was estimated by the Updegraff method [25]. In the first step, alcohol insoluble residue (AIR) was prepared using ethanol. This AIR was further mixed with the acid nitric reagent and then put in a water bath at boiling temperature for 30 min. It was washed with Milli-Q and then ethanol until the cellulose appears. This residue was then kept in the oven at 37 °C overnight. The next day, 67% H_2SO_4 was added to it and put on a shaker for the complete dissolution of cellulose. The cellulose estimation of the sample was done by anthrone reagent [26].

2.3.3 Determination of hemicellulose

Estimation of the hemicellulose content was done by fibre analysis method [27]. Difference between neutral detergent fibre (NDF) and acid detergent fibre (ADF) gives the total hemicellulose in a given sample. All the chemicals used in this experiment were procured from Fisher Scientific Ltd. One gramme sample is treated with 100 mL Neutral detergent solution with the addition of 2 mL decahydronaphthalene and 0.5 g sodium sulfite in refluxing flask for 1 h. The sample was then transferred to the pre-heated crucible. The sample was filtered and washed two times with Milli-Q and acetone consecutively. The sample was transferred to the hot air oven at 100 °C for 8 h and then put in a desiccator. On the other hand, for the determination of ADF, a 1 g sample is treated with an acid detergent solution in a refluxing flask for 1 h and the same procedure was followed as for NDF [28].

$$\text{Hemicellulose } (\%) = \text{NDF} - \text{ADF}$$

2.3.4 Determination of lignin content

This was performed by the NREL protocols. Before the estimation of the lignin, the extraction of the sample was performed using Milli-Q in 1:20 ratio [29]. Three hundred milligrammes extracted sample with 72% H_2SO_4 in a pressure tube was put in a water bath at 30 °C for 1 h with periodic stirring. Four percent dilution was done using Milli-Q, and the tube was autoclaved for 1 h. Pre-heated and labelled crucible was weighted, and the sample was filtered by a filtration unit in the crucible and then washed twice. After that, the sample was dried at 105 °C for 4 h and cooled in a desiccator, and acid insoluble residue (AIR) was calculated. Later, the crucible was transferred to a muffle furnace at 575 °C for 24 h to calculate the percentage of ash in the filter [30]. The lignin content (%) was

calculated using AIR (%)and ash in the filter (%). Acid-soluble lignin was calculated by measuring absorbance at 280 nm [31]. All the calculations were performed using the formulas mentioned below.

$$\%\ \text{AIR} = \frac{(\text{crucible plus AIR weight} - \text{crucible weight})}{\text{sample loading in crucible}} \times 100$$

$$\%\ \text{Ash in filter} = \frac{(\text{crucible plus ash weight} - \text{crucible weight})}{\text{sample weight}} \times 100$$

$$\%\ \text{Acid insoluble lignin} = \%\ \text{AIR} - \%\ \text{ash in filter}$$

$$\%\ \text{Acid} - \text{soluble lignin} = \frac{\text{UV}_{\text{absorbance (280nm)}} \times \text{volume of filtrate} \times \text{dilution}}{\text{absorptivity} \times \text{dry weight} \times \text{pathlength}}$$

2.3.5 Determination of starch

Starch was estimated by anthrone reagent. Chemical used in this composition was anthrone reagent which was prepared in 95% sulphuric acid, 52% perchloric acid and 80% ethanol. Initial step of this was done by washing the sample with 80% ethanol repeatedly till the colour of the residue was removed. Then, the residue was dried in a hot air oven, mixed with perchloric acid and incubated for 20 min. The tubes were centrifuged at 10,000 rpm for 10 min, and the supernatant was taken out. The glucose standard was prepared using 1 mg/mL stock solution. Anthrone is added to the supernatant of the sample and glucose standard and heated at boiling temperature. After heating, the reading was taken on a spectrophotometer at 630 nm for colour detection, and starch concentration was calculated [32].

2.3.6 Fourier-transform infrared (FTIR) spectroscopy

The changes in functional groups after providing the acidic pre-treatment were analysed by FTIR spectroscopy (PerkinElmer 400 FTIR/FTIR) control with a frequency range of 4000–400 cm^{-1} with the control sample. The pellet preparation was done with potassium bromide (KBr), and scanning was performed at 4 cm^{-1} resolution [33].

2.3.7 Thermogravimetric analysis (TGA)

The TGA of raw and pre-treated samples was performed by PerkinElmer TGA 4000. Samples were heated from room temperature to 600 °C at a heating range of 10 °C min^{-1} with 10 mL/min nitrogen flow rate [33, 34].

2.3.8 Scanning electron microscopy (SEM)

The changes in the morphology of the pre-treated sample were investigated with SEM (Model: EV018 Zeiss, Germany) with 5 kV voltage at × 10,000 magnifications. The breakdown of the cell wall was identified after SEM analysis [34].

2.4 Response surface methodology (RSM) experimental design

In this study, three independent variables (acid concentration (X_1, vol.%), time (X_2, min) and temperature (X_3, °C) with two dependent variables (reduced sugar (mg/mL) and cellulose (%))) as response were chosen for experiment design. These three independent variables along with their respective ranges were found to be critical parameters for maximum sugar production. It is expected in the design that the independent variable is managed by experiments with minor errors. The main aim of this model was to optimize the response variables (Y). This statistical model gives the approximate correlation between independent variables and dependent variables (responses) [35]. Experiments were run in random order for negligible error in the model. This empirical model is formed by the responses which show the correlation of independent variables with the help of a polynomial equation:

$$\text{Y} = \beta_o + \sum_{i=1}^{k} \beta_i \text{X}_i + \sum_{i=1}^{k} \beta_{ii} \text{X}_i^2 + \sum_{i<j}^{k} \beta_{ij} \text{X}_i \text{X}_j + e$$

where Y is the variable of response, β_o is the constant coefficient,$\beta_{\rangle}$ is the linear constant, $\beta_{\rangle\rangle}$ is the quadratic coefficients, $\beta_{\rangle|}$ is the interaction effect of coefficients, $\text{X}_{\rangle}$ and $\text{X}_{|}$ are the coded values used for variable parameters and ⌉ is the random error [36].

2.4.1 Design of pre-treatment experiments

Acidic pre-treatment was performed with sulphuric acid. The experiment design was done to optimize the conditions and determine the maximum reduced sugar liberation that was further used in ethanol production. Biomass sample was mixed with acid with a 10% biomass loading. Pre-treatment was done in an autoclave and neutralization of the sample is done by the $CaCO_3$ for the detoxification. After pre-treatment, the estimation of reduced sugar is done by the DNSA method, and cellulose is determined by the Updegraff method. Experiment design for pre-treatment was done by the central composition design (CCD), and the response surface methodology (RSM) approach was used to

Table 1 Composition analysis of *Pistia stratiotes*

Pistia stratiotes		
S. No	Biochemical composition	(%) of component
1	Total solid	92.85 ± 0.11
2	Ash content	18.36 ± 1.31
3	Moisture content	7.15 ± 0.11
4	Cellulose	25.90 ± 0.017
5	Hemicellulose	18.44 ± 0.71
6	Lignin	25.25 ± 1.15
7	Starch	0.6 ± 0.03

analyse the data [22]. Three variables, including temperature (93.18 °C, 100 °C, 110 °C, 120 °C and 126.82 °C), time (12.96, 30, 55, 80 and 97.04 min) and acid concentration (0.15%, 0.75%, 1.63%, 2.5% and 3.1%), were selected to find out the optimized condition. Reducing sugar and cellulose were recorded as a response for this design. A total of 20 runs in design with three levels (−1, 0, + 1) with a 1.68 alpha value were performed.

After the pre-treatment, the detoxification of liquid residue was done by calcium carbonate ($CaCO_3$) powder. Toxic compounds may be retained in the hydrolysate, which can inhibit the microorganism's growth. After pre-treatment in an autoclave, the residue was centrifuged, and the liquid hydrolysate was separated from the solid fraction. This liquid fraction was detoxified by adding calcium carbonate to liquid hydrolysate, and pH was monitored by the amount of calcium carbonate addition [37].

2.5 Total reducing sugar analysis

Total reducing sugars in the liquid hydrolysate were determined by the 3,5-dinitrosalicylic acid (DNSA) method [38]. DNSA reagent was prepared and stored in an amber bottle for further use. Glucose stock solution was prepared for standard curve determination. This DNSA reagent was added to the standards as well as a sugar sample, boiled for 15 min and cooled down at room temperature. The samples' optical density (OD) was measured at 540 nm for the estimation of reducing sugar with the help of a standard curve [38]. The total reducing sugar of raw and pre-treated samples was quantified.

2.6 Fermentative process

2.6.1 Microorganisms

Saccharomyces cerevisiae (ATCC 834) and *Pichia Stipitis* (NCIM 3594) for ethanol production were procured from the National Collection of Industrial Microorganism (NCIM), National Chemical Laboratory, Pune (India) in the slant

Table 2 Experimental design with three independent factors (time, temperature and acid conc.) with cellulose and sugar (dependent factors)

Experiments	Acid conc. (%)	Time (min)	Temperature (°C)	Cellulose (%)	Sugar (%)
1	3.1	55	110	19.81 ± 0.47	24.01 ± 3.87
2	0.75	30	120	10.88 ± 1.35	10.36 ± 1.21
3	2.5	80	120	24.55 ± 1.98	19.86 ± 0.0
4	1.63	12.96	110	14.46 ± 0.43	12.75 ± 0.74
5	2.5	30	120	22 ± 3.14	19.88 ± 1.31
6	1.63	55	110	16.02 ± 1.22	15.89 ± 0.34
7	0.75	80	100	9.31 ± 0.58	6.76 ± 0.13
8	1.63	55	110	17.7 ± 1.27	14.99 ± 0.27
9	2.5	30	100	12.84 ± 1.38	13.54 ± 3.82
10	1.63	55	126.82	17.6 ± 1.37	17.02 ± 0.89
11	1.63	55	110	14.9 ± 4.17	16.49 ± 0.52
12	1.63	55	93.18	10.93 ± 2.02	4.47 ± 0.16
13	1.63	55	110	15.56 ± 1.80	14.72 ± 0.28
14	0.75	80	120	14.1 ± 1.51	8.53 ± 0.35
15	1.63	55	110	16.63 ± 0.25	15.33 ± 0.18
16	1.63	97.04	110	15.2 ± 1.25	11.79 ± 0.10
17	2.5	80	100	15.17 ± 0.86	11.22 ± 0.15
18	0.75	30	100	8.45 ± 1.19	4.1 ± 0.11
19	1.63	55	110	16.79 ± 2.48	15.75 ± 0.50
20	0.15	55	110	4.91 ± 0.67	3.07 ± 0.02

form (active culture) at 4 °C. Both of the strains were maintained in MGYP media Petri plates.

2.6.2 Inoculum preparation

S. cerevisiae and *P. stipitis* were grown in the MGYP media (10 g/L glucose, 3 g/L malt extract, 3 g/L yeast extract and 5 g/L peptone) [39]. Seed culture was prepared in the conical flask with 50 mL of media, inoculated with a loopful of cells and incubated in an orbital shaker with 150 rpm at 30 °C. The mid-exponential phase culture was used to inoculate culture media with an optical density of around 0.8, which was measured at 600 nm in a UV–vis spectrophotometer [40]. The fermentative media were inoculated with 5% (v/v) of seed culture for ethanol production.

2.6.3 Fermentation for ethanol production

Fermentation was performed in a 250-mL shake flask with 150 mL working volume containing synthetic media (malt extract 0.3%, yeast extract 0.3%, peptone 0.5% and glucose 1%) and hydrolysate detoxified by $CaCO_3$ in triplet. For fermentation, 5% (v/v) of the seed culture was inoculated in detoxified hydrolysate. All the experiments were performed for 56 h. Sample collection was done periodically for the determination of cell growth, sugar consumption and ethanol production. The ethanol was separated by gas chromatography (Clarus 580, PerkinElmer, USA) equipped with a ZB-wax column (60 m × 0.32 mm internal diameter 0.25 µm; Phenomenex, UK). Oven temperature was set in between 150 and 180 °C. Ethanol data was captured from programmed software (Total Chrome Workstation Ver 6.3) that was pre-installed. The ethanol estimation in the sample was done by comparing it with the standard of ethanol (EMSURE ACS, Sigma-Aldrich, USA). All samples are performed in triplicates, and the concentration was mentioned in the grammes per litre [41].

Ethanol-related kinetic parameters The kinetic parameters related to ethanol yield were analysed based on the aforementioned reports [42–44].

$$\text{Ethanol yield(YE)} = \frac{\text{ethanol concentration}\left(\frac{\text{g}}{\text{L}}\right)\text{in fermentation media(Ef)} \times 1}{\text{sugar consumed } \left(\frac{\text{g}}{\text{L}}\right)}$$

$$\text{Volumetric ethanol productivity}\left(\frac{\frac{\text{g}}{\text{L}}}{\text{h}}\right) = \frac{\text{Ehanol concentation}\left(\frac{\text{g}}{\text{L}}\right)\text{in fermented broth}}{\text{fermentation time (h)}}$$

$$\text{Fermentation efficiency}(\%) = \frac{\text{ethanolyield(YE)} \times 100}{\text{theoretical ethanol yield}}$$

3 Result and discussion

3.1 Composition analysis of raw sample

In the *P. stratiotes*, the carbohydrate concentration is high so that it can be used as a carbon source for ethanol production (Table 1). The amount of cellulose is observed as 25.29% with a low level of lignin (16.73%) in comparison to another study by Sutaryo et al. The lignin composition is 34.85% [45]. In the present study, 18.44% hemicellulose was observed, while Mishima et al. reported 17.3% [46]. These variations in compositions could be due to cultivation conditions, age and location. Approximately 60% of biomass was composed of carbohydrates in this *P. stratiotes*. *P. stipitis* utilizes both pentose and hexose sugar from the hydrolysate by which maximum production of ethanol may be achieved. Considering the convenience and utilization, this plant was taken as a carbon source for bioethanol production.

3.2 Characteristics of pre-treated *Pistia stratiotes*

3.2.1 Acidic pre-treatment

Sulphuric acid pre-treatment has a high impact on the biomass for depolymerization of cellulose and hemicellulose with a maximum amount of sugar found, which is further utilized in the fermentation process for ethanol production. After performing all the experiments, maximum sugar was found at 3.1% acid concentration. After performing the optimization condition, range of sugar content was 3.07 to 24.01%, and the target was to find out 23% sugar content.

Table 3 Analysis of variance for coded coefficients

Source	DF	Adj SS	Adj MS	*F*-value	*P*-value
Model	8	570.888	71.361	34.48	0.000
Linear	3	501.680	167.227	80.80	0.000
Acid conc	1	358.453	358.453	173.20	0.000
Time	1	0.715	0.715	0.35	0.569
Temp	1	142.512	142.512	68.86	0.000
Square	3	61.914	20.638	9.97	0.002
Acid conc.*acid conc	1	8.007	8.007	3.87	0.075
Time*time	1	20.559	20.559	9.93	0.009
Temp.*Temp	1	43.310	43.310	20.93	0.001
2-way interaction	2	7.294	3.647	1.76	0.217
Acid conc.*time	1	1.256	1.256	0.61	0.452
Acid conc.*Temp	1	6.038	6.038	2.92	0.116
Error	11	22.765	2.070		
Lack of fit	6	20.678	3.446	8.26	0.018
Pure error	5	2.087	0.417		
Total	19	593.653			

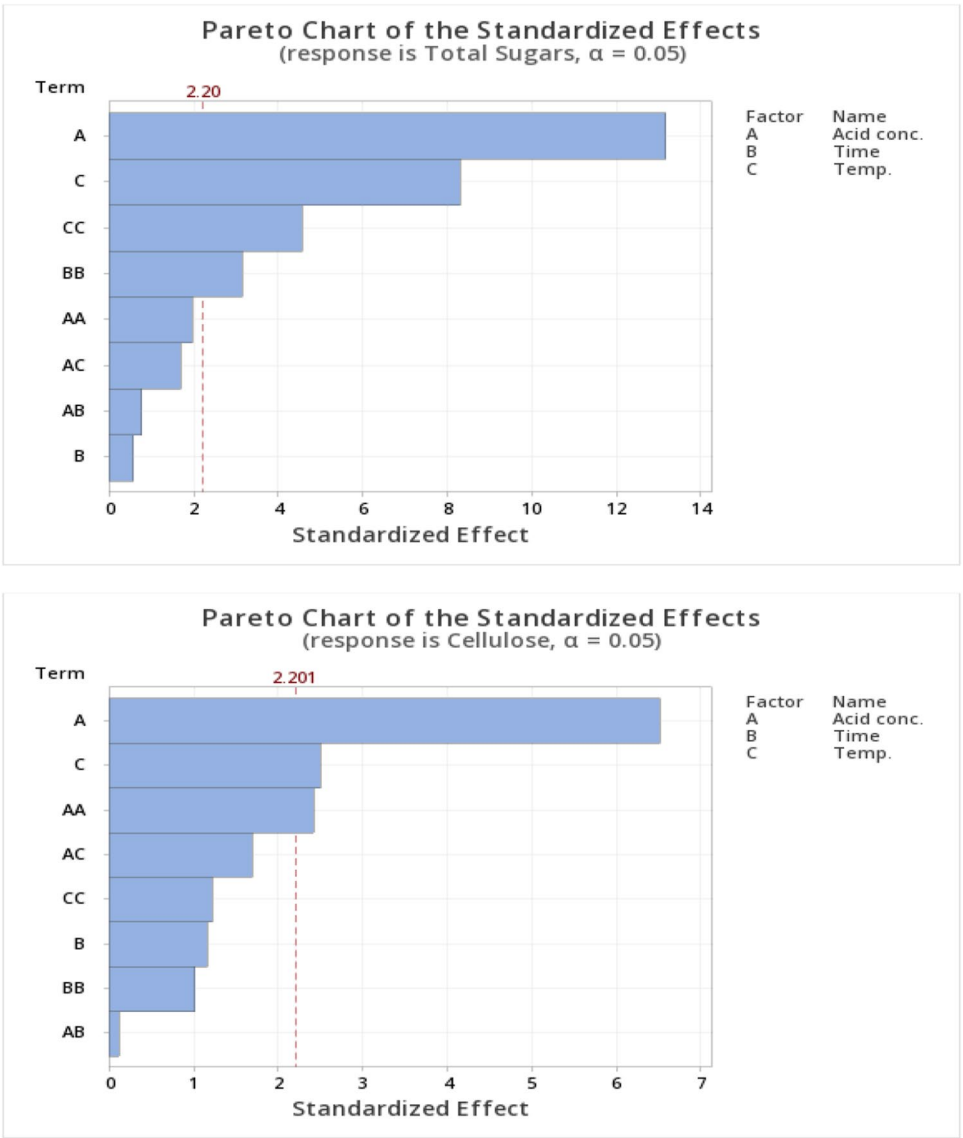


Fig. 1 Pareto chart shows standardized effect of independent factors on sugar and cellulose. Parameters of independent variable combination crossing the reference line shown by dotted red line are statistically significant

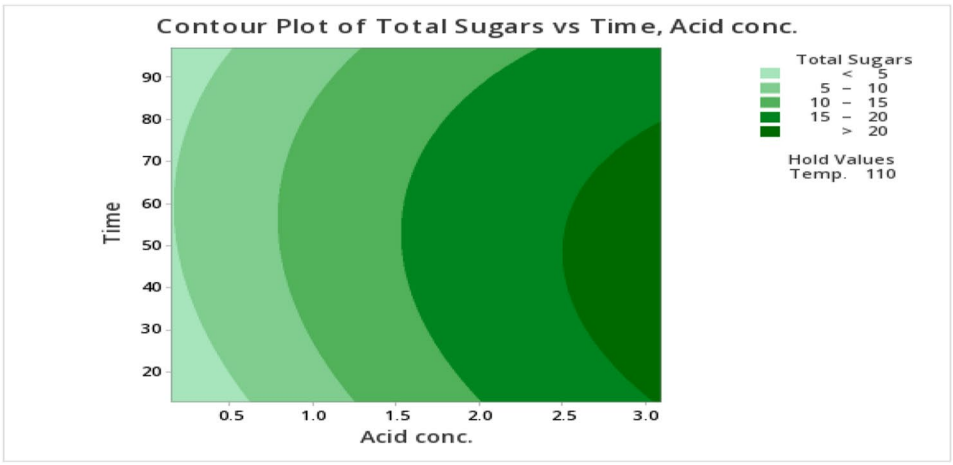
Similarly, range of cellulose content was 4.91 to 24.55%. After the optimization, the best condition was achieved as 2.5% sulphuric acid, 120 °C with a duration of 15 min. by autoclave. The outcome of this experiment was 23.44% sugar with 68% hydrolysate.

All the details of independence and responses from each experiment in experimental design are mentioned in Table 2.

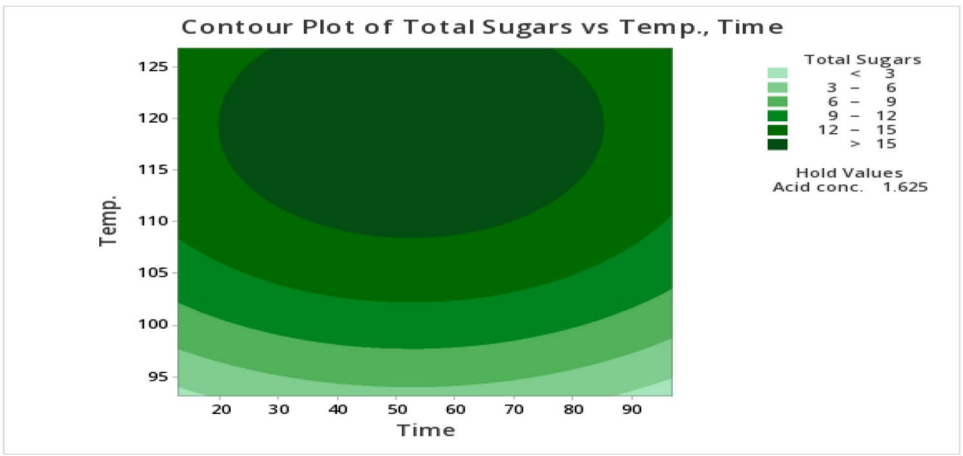
3.2.2 RSM model for pre-treatment

The coefficient of determination and statistical significance were estimated in the current study. The significance and analysis of variance for coded coefficients are represented in Table 3. The model for optimization is significantly based on the *P*-value. Model terms acid concentration, temp., square, time*time and temp.*temp were most significant in the optimization of sugar ($P < 0.05$). This regression model is performed with an alpha value of 1.68.

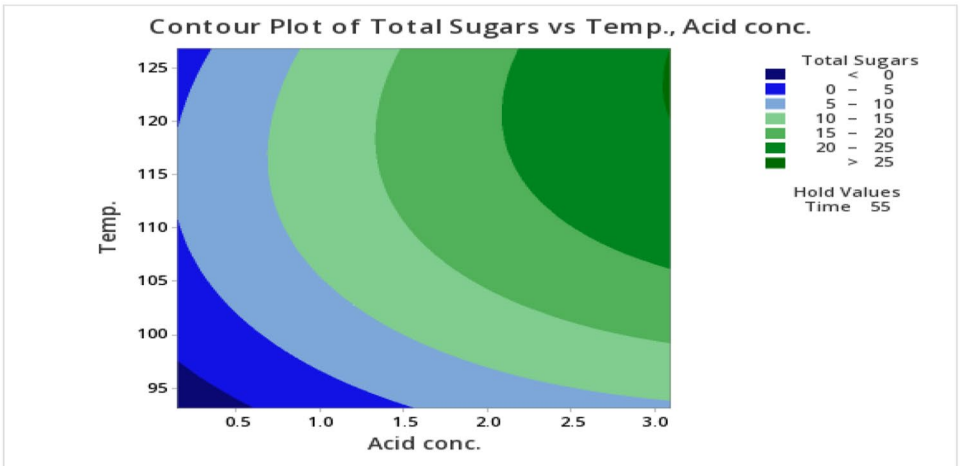
The R^2 (coefficient of determination) values are used to determine the goodness of fit of this statistical optimization model. In this model, the lack-of-fit *F*-value is 8.26 for optimization of the dependent variable (i.e. sugar), which is non-significant. This non-significance of lack of fit suggests that our model is reliable. The regression model in this study gives a regression equation that determines the relationship between the response (i.e. sugar) and the independent variable uncoded values (acid con., temp. and time). The regression equation is shown in Eq. 1.



(a)



(b)



(c)

Fig. 2 Contour plot shows the detailed information of independent and dependent variable correlation. **a** Total sugar vs. time and acid concentration at a constant temperature. **b** Total sugar vs. temp. and time at constant acid conc. **c** Total sugar vs. temp. and acid conc. at constant time

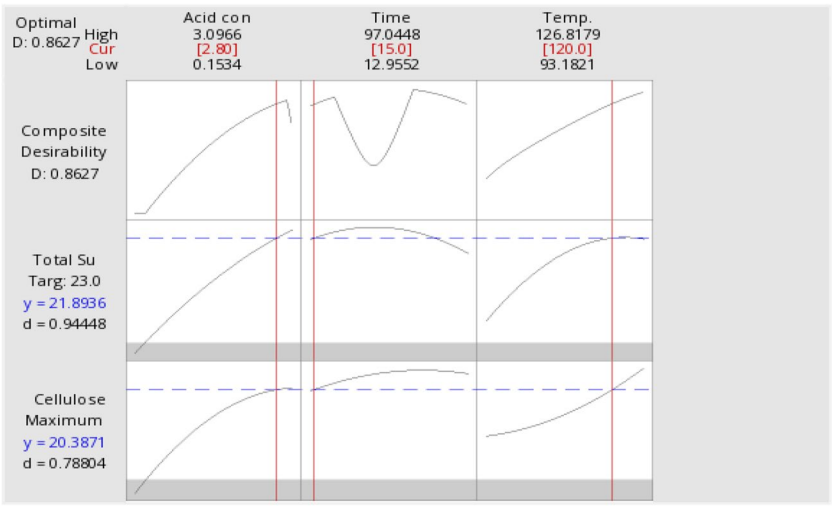


Fig. 3 Optimized condition after performing RSM shows that the composite desirability score is 0.8627 which is good definition that both responses are very close to the entire things in the model. Sugar has high desirability as compared to cellulose shown by the score of 0.94448 for sugar and 0.78804 for cellulose. Outcome of the result find out with 2.8% acid at 120 °C with 15 min

$$\begin{aligned}\text{Total sugars} = &-231.0 - (0.91 \times \text{acid conc.}) + (0.2305 \times \text{time})\\ &+(3.976 \times \text{ Temp.}) - (0.974 \times \text{acid conc.} \times \text{ acid conc.})\\ &-(0.001911 \times \text{time} \times \text{time}) - (0.01734 \times \text{Temp.} \times \text{ Temp.})\\ &-(0.0181 \times \text{acid conc.} \times \text{ time}) + (0.0993 \times \text{acid conc.} \times \text{ Temp.})\end{aligned} \quad (1)$$

Equation no. 1: Regression equation in uncoded units R^2 value is 96.17% (*R*-squared is a statistical measurement of data which shows how close the data is to the regression line) with 93.38% adjusted R^2 value. From R^2 and R^2 (adj.) values, R^2 (pred.) predicted value was confirmed at 80.34% which indicates the regression model predicts the responses in new findings very well.

3.2.3 Pareto chart

The Pareto chart determines the significance of the effect of independent variables on total sugar and cellulose. This

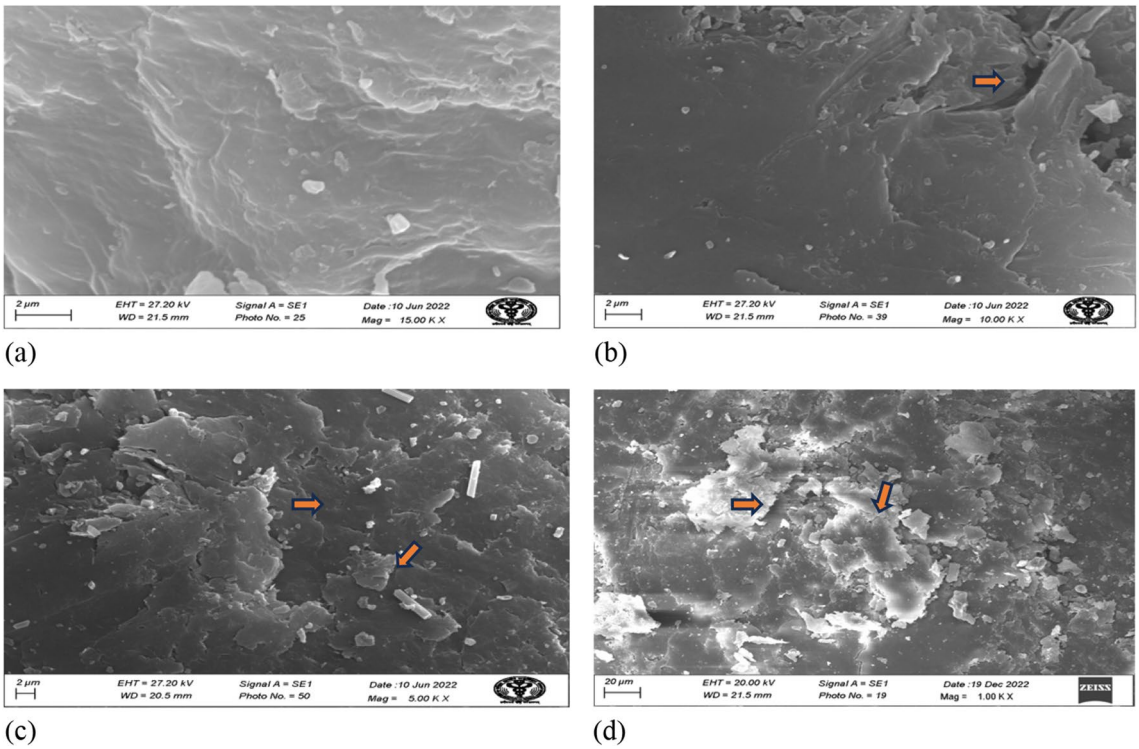


Fig. 4 SEM images of **a** raw biomass, **b** harsh pre-treatment conditions (2.5% acid, 80 min. with 120 °C), **c** mild pre-treatment conditions (0.75% acid, 30 min. with 100 °C) and **d** optimized condition

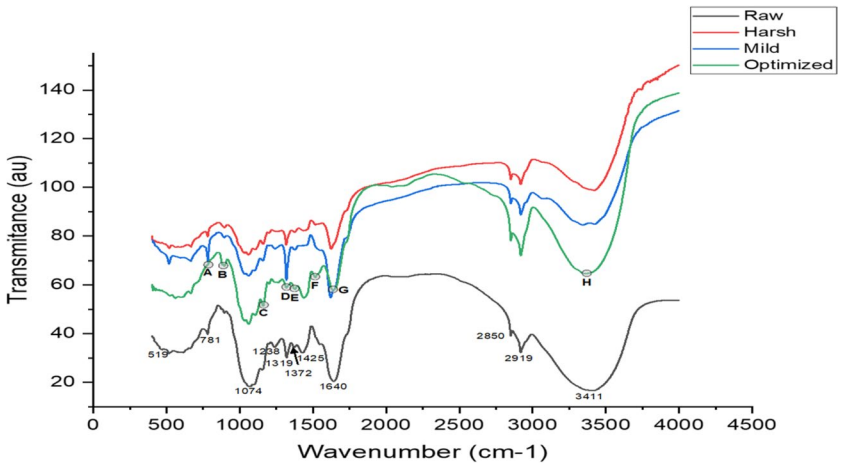


Fig. 5 FTIR spectra of raw, mild (0.75% acid, 30 min, 100 °C), harsh (2.5% acid, 80 min, 120 °C) and optimized pre-treated sample. These spectra show the changes in the functional groups after pre-treatment (A–H) denotes the changes in the peaks which shows the breakdown of the cellulose and removal of lignin

Pareto chart gives all the details of the independent variable in pictorial form. In the Pareto chart of the standardized effect for sugar, the reference line was at 2.20, whereas for cellulose, it was at 2.201. the parameters crossing the reference line were statistically significant. The outcome of the chart determines that the acid conc. (A), temperature (C), temperature*temperature (CC) and time*time (BB) are statistically significant at 0.05 level. On the other hand, significant effect of the independent variable on the cellulose shown in chart no. 2 in Fig. 1 shows acid conc. (A), temperature (C) and acid conc.*acid conc. (AA) are statistically significant. This statistical significance of the independent variable proves that this model is valid for the optimization of pre-treatment.

3.2.4 Contour plot

Contour plot represents the two-dimensional (2D) image of factors and responses and shows the relationship between the independent variable and dependent variables (Fig. 2). While keeping the temperature constant, sugar (> 20%) was obtained at a combination of acid concentration (2.5–2.6%) and time (0–75 min) (Fig. 2a). While keeping the acid conc. constant, sugar conc. (> 15%) was obtained at the temperature (110–125 °C) with time (20–85 min) (Fig. 2b). Similarly, while keeping the time constant, sugar conc. (20–25%) was obtained at acid concentration (2.2–3.0%) and temperature (107–125 °C) (Fig. 2c).

3.2.5 Optimization plot

After performing the optimization of acidic pre-treatment, an optimized condition was found in which independent variables are acid conc. 2.8%, time 15 min with 120 °C. This condition is performed separately to find out the reliability of the condition. After performing the optimized condition, maximum sugar was 23.44% in the hydrolysate which shows the reliability of the model. Minitab software creates a model in which independent variables are added to find the target response. Optimization is calculated by combined desirability ranges from 0–1 [47, 48]. For the optimization experiment, the separate desirability for the entire data is shown in Fig. 3. Sugar has an independent desirability result of 0.94448 as the predicted feedback of 21.8936 which is very close to the target of 23. Cellulose has a moderate desirability outcome that is 0.78804 as the predicted outcome of 20.3871 which is less than the target of 24.55 as mentioned in the parameter table. The composite desirability of 0.8627 is a good score which represents that both the responses were very close in their absolute settings. The optimization of responses did not give an ideal composite desirability score as the sugar and cellulose contents were not in absolute settings but in a justifiable range.

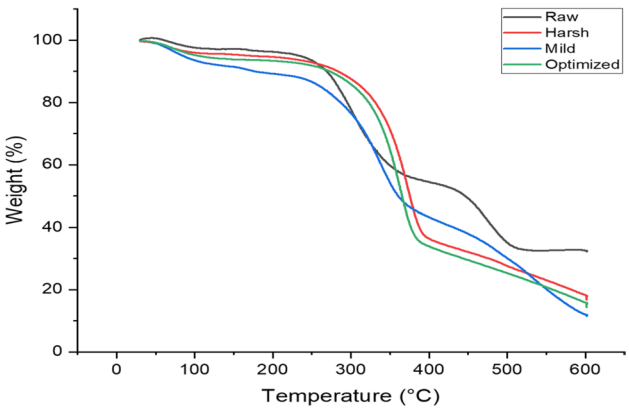


Fig. 6 Thermogravimetric analysis of raw (untreated), mild (0.75% acid, 30 min, 100 °C), harsh (2.5% acid, 80 min, 120 °C) and optimized pre-treated samples confirm the degradation of the compounds in the material or lose the stability of material at a particular temperature optimized condition have high degradation of cellulose and lignin

Table 4 Kinetic parameters for ethanol production by *S. cerevisiae* and *P. stipitis* stains with synthetic media and hydrolysate

Kinetic parameters	*Saccharomyces cerevisiae* in Synthetic media	*Saccharomyces cerevisiae* in hydrolysate	*Pichia stipitis* in synthetic media	*Pichia stipitis* in hydrolysate
Initial total sugar(g/L)	10.04 ± 0.29	10.04 ± 1.36	9.97 ± 0.76	9.99 ± 0.21
Maximum ethanol conc.(g/L)	3.32 ± 0.96	3.26 ± 0.71	3.57 ± 0.31	3.57 ± 2.02
Maximum time (hours)	30	36	32	36
Maximum sugar consumed (%)	97.6 ± 0.29	85.9 ± 1.34	94.9 ± 0.74	87.9 ± 0.21
Ethanol productivity(g/L/h)	0.11 ± 0.04	0.09 ± 0.02	0.10 ± 0.01	0.09 ± 0.06
Growth rate(g cells/L/h)	0.38 ± 0.01	0.30 ± 0.004	0.68 ± 0.03	0.56 ± 0.01
Maximum sugar consumption rate(g/L/h)	0.29 ± 0.01	0.24 ± 0.002	0.24 ± 0.001	0.24 ± 0.00
Ethanol yield coefficient (g/g)	0.39 ± 0.02	0.37 ± 0.06	0.41 ± 0.03	0.39 ± 0.01

The conditions found after this RSM model are mentioned in Fig. 3.

3.3 Characterization of raw and pre-treated biomass

3.3.1 Morphological changes in biomass (SEM analysis)

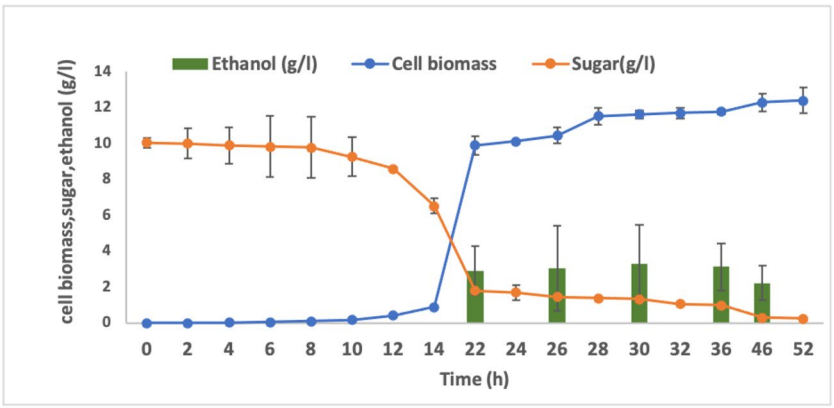
The morphology of *P. stratiotes* changes after acidic pre-treatment. It is more destructive after providing acidic pre-treatment to the sample as compared to the raw material. Maximum disruption was observed in optimized acidic conditions which is denoted by Fig. 4d. However, harsh conditions (2.5% acid conc.) also show a high amount of disruptions denoted by Fig. 4b. Mild condition (0.75% acid conc.) in pre-treatment is shown by Fig. 4c. These disruptions are due to the effect of temperature and acid by which smooth surface of the biomass is converted into the rough surface which shows that the bonds between cellulose and hemicellulose are broken and removal of lignin from the sample increases the surface area of the biomass to increase the accessibility to the enzymes for saccharification. This type of structural change is also seen in the case of elephant grasses when treated with sulphuric acid [49].

3.3.2 FTIR spectrum of raw sample and pre-treated sample

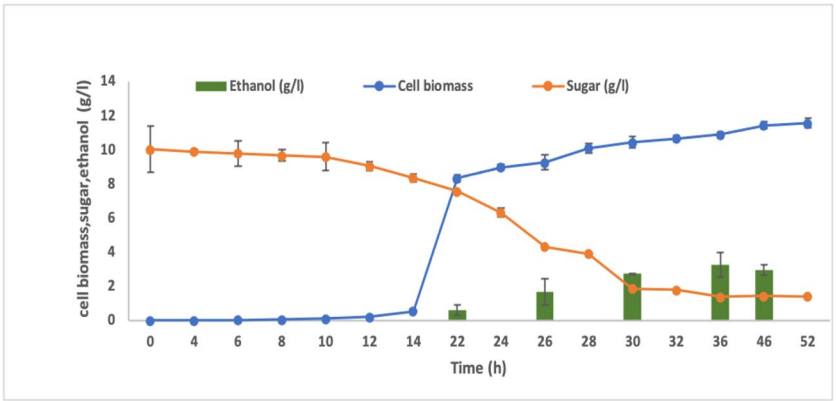
The FTIR spectrum provides detailed information regarding the functional group and its properties. The FTIR spectra of raw, mild (0.75% acid, 30 min, 100 °C), harsh (2.5% acid, 80 min, 120 °C) and optimized pre-treated samples of *P. stratiotes* (Fig. 5) showed a broad peak (H) around 3500–3200 cm^{-1} in the optimized condition, which shows N–H starching of primary amines and free O–H stretching of the OH group present in cellulose [50]. By acidic pre-treatment, these O–H bands were reduced [36]. The peak between 2848 and 2851 cm^{-1} in each pre-treated and raw sample represents the characteristic C–H stretching, which represents the vibration of cellulose [51]. The peak observed at 1640 cm^{-1} (G) shows water absorption, and in the optimized pre-treated condition, the intensity of the peak decreased due to aromatic skeletal vibration giving information regarding lignin removal [52]. In the pre-treated condition, the peak at 1515 cm^{-1} (F) relates to the aromatic skeletal stretching. Acidic pre-treatment reduced the adsorption of bands, but acid-soluble lignin occurred in the solid fraction of the pre-treated sample. The intensity of the peak at 1374 cm^{-1} (E) increased due to breakage in ether groups present in lignin [53]. As compared to the pre-treated sample, the peak at 1316 cm^{-1} (D) in the raw sample is more prominent due to the C–H ring vibration, indicating the presence of cellulose. CH_2 bending, which represents the cellulose with skeletal vibration of the C–C bond, shows the presence of cellulose [54]. The peak between 1200–1000 cm^{-1} (C) displays the presence of cellulose and hemicellulose in the sample with C–O starching and C–H rocking vibration correlating with cellulose structure [55]. The range of the aromatic band is 950–700 cm^{-1} corresponding to the β-glycosidic linkage between sugar units of cellulose and hemicellulose [15]. FTIR indicates the effective removal of other contents apart from cellulose by acidic treatment of *P. stratiotes*. The appearance of a peak at 894 cm^{-1} (B), an absorption peak for cellulose, in the pre-treated sample shows the C–O–C stretching in β-glycosidic linkage in cellulose and hemicellulose [56]. The peak intensity at 781 cm^{-1} (A) is decreased in the pre-treated sample as compared to the raw sample, showing the removal of lignin by the acidic treatment [57].

3.3.3 Thermogravimetric analysis (TGA)

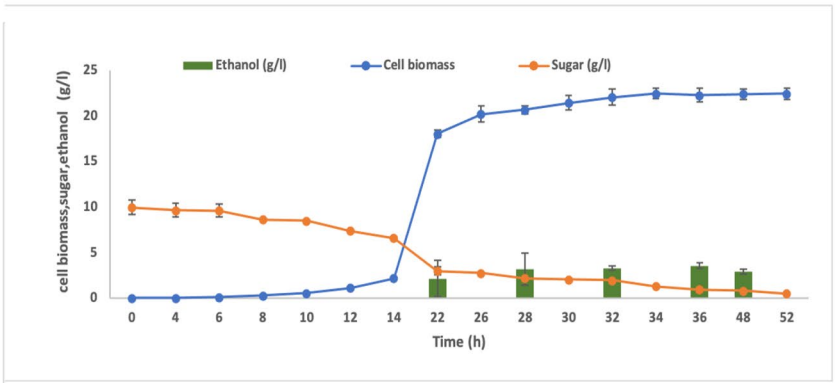
Raw sample, mild (0.75% acid and 30 min at 100 °C), harsh (2.5% acid and 80 min at 120 °C) and optimized pre-treated sample (acid conc. 2.5% and 15 min. at 120 °C) were thermogravimetrically analysed to differentiate their degradation characteristics. *P. stratiotes* biomass contains four types of weight which can be degraded (moisture,



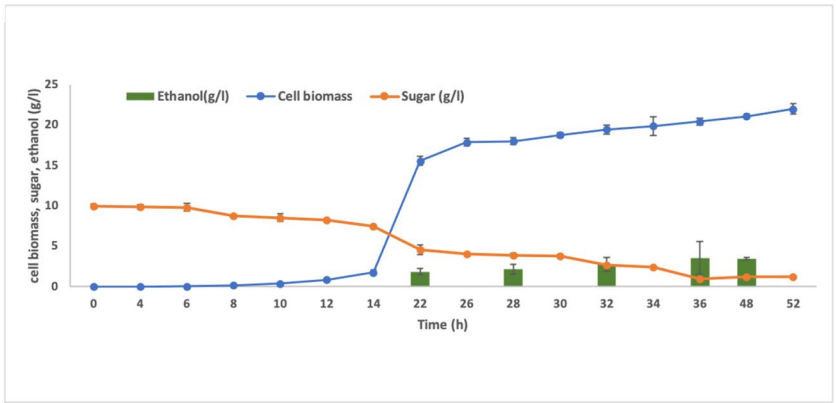
(a)



(b)



(c)



(d)

Fig. 7 The outcome of the fermentation process in four different conditions, using *Saccharomyces cerevisiae* **a** with synthetic sugar and **b** with hydrolysate and by using *Pichia stipitis* **c** with synthetic sugar and **d** with hydrolysate, was done by calculation of cell biomass, sugar consumption and ethanol production

Table 5 Comparison of different lignocellulosic biomass capability of conversion into ethanol

Biomass	Strains	Ethanol yield (g/g)	Reference
Lignocellulosic biomass	Engineered *Saccharomyces cerevisiae*	0.43	[61]
Water hyacinth	*S. cerevisiae*, *P. stipitis* and *C. lusitaniae*	0.32, 0.44, 0.49	[59]
Rubberwood waste	S. cerevisiae	0.14	[62]
Sugarcane bagasse	*S. cerevisiae*	0.41	[63]
Kans grass	*S. cerevisiae*	0.46	[64]
Pistia stratiotes	*S. cerevisiae* and *P. stipitis*	0.37 and 0.39	Present study

cellulose, hemicellulose and lignin) during the TGA shown in Fig. 5 [7]. In all samples, initial weight loss occurred at 100 °C due to evaporation of water. Temperature from 200–300 °C leads the way to depolymerization of hemicellulose and breakage of glycosidic bonds in cellulose. Unexpected loss of weight in mild conditions and untreated raw material at 250 °C is due to the degradation of hemicellulose. This was primarily degraded due to the presence of acetyl group as compared to cellulose and lignin. Lignin degradation occurs simultaneously at the temperature range 200–500 °C and range 250–400 °C is confirmed for cellulose degradation. Maximum weight loss of the pre-treated sample of approximately 53% occurs in the range of 250–400 °C due to degradation of cellulose. It is much higher as compared to the raw sample [34, 58], and on the other hand, optimized and harsh condition degradation goes simultaneously as the same pattern. Degradation in the optimized condition is seen maximum in Fig. 6.

3.4 Fermentation by Saccharomyces cerevisiae and Pichia stipitis

3.4.1 Analysis of ethanol by gas chromatography (GC-FID)

The retention time for the standard of ethanol was 3.61 min with a complete 10-min cycle. *S. cerevisiae* used in fermentation have the ability to utilize hexose for ethanol production, while *P. stipitis* has the capability for the usage of pentose and hexose for ethanol production. The fermentation was done with *S. cerevisiae* NCIM 3594 and *P. stipitis* NCIM 3498 separately to find out the maximum outcomes of the fermentation process. Fermentation done by *S. cerevisiae* and *P. stipitis* utilizes all sugar within 52 h. *P. stratiotes* hydrolysate after pre-treatment contained a maximum of 23.44% sugar, which was further used in the process of fermentation by *S. cerevisiae* and *P. stipitis*. Maximum glucose was consumed within 48 h and converted into ethanol. Starting sugar was taken at 10.04 g/L. The fermentation condition was carried out at 30 °C at 150 rpm. In the fermentation maximum, 86% sugar was consumed from the hydrolysate by the *S. cerevisiae*, while 88% sugar was consumed by *P. stipitis*. During the initial stage of fermentation, intake of sugar by the strains was slower, and cell growth increases after 14 h when sugar consumption increased. Maximum ethanol concentration was found by the *S. cerevisiae* in synthetic media in 30 h, while in hydrolysate, it comes in 36 h (3.25 g/L) with 0.37 g/g ethanol yield. On the other hand, *P. stipitis* strain utilizes both pentose and hexose sugars and the maximum ethanol concentration in synthetic media comes in 32 h, while in hydrolysate, it comes in 36 h (3.57 g/L) with 0.39 g/g ethanol yield (Table 4 and Fig. 7). The ethanol concentration in the hydrolysate was approximately similar the synthetic media because of detoxification of hydrolysate which removes high amount of toxic compound from the hydrolysate by which production of ethanol increased. In the Sunwoo et al. [59] (Table 5) study, *S. cerevisiae* gives less ethanol yield (0.32 g/g) as compared to the present study (0.37 g/g), and *P. stipitis* gives higher production (0.39 g/g) as compared to *S. cerevisiae*. This determines that both the stains give a very good amount of ethanol by this aquatic weed. The kinetic parameters of our study are comparable just like another study on sugarcane molasses [60].

4 Conclusions

P. stratiotes is used as a biomass for ethanol production because of its high carbohydrate content with a maximum amount of cellulose. Acidic pre-treatment enhances sugar content accessible to the microbes in the sample. In previous studies that used *P. stratiotes*, the optimization of acidic pre-treatment was not performed. Thus, we used the response surface methodology (RSM) analysis using Minitab statistical software to optimize acidic pre-treatment conditions to obtain maximum sugar that is affected by three independent variables, viz. acid concentration, time and temperature. The optimized condition (2.5% acid conc. at 15 min and 120 °C) was used to obtain hydrolysate with 23.44% reduced sugar, which was further detoxified with calcium carbonate and subsequently used for ethanol production. *S. cerevisiae* can only use hexose sugar for fermentation, while *P. stipitis* can use both pentose and hexose sugar. *S. cerevisiae* produced

ethanol conc. 3.32 g/L with 0.39 g/g ethanol yield from the synthetic media, and in the hydrolysate, ethanol conc. 3.25 g/L with 0.37 g/g ethanol yield was found. *P. stipitis* produced 3.57 g/L ethanol conc. with 0.41 g/g ethanol yield from the synthetic media, while in hydrolysate, 0.39 g/g ethanol yield was calculated. Bioethanol from lignocellulosic biomass is a sustainable approach as it is a requirement to replace fossil fuels.

Acknowledgements We would like to thank the senior management of Delhi Technological University (DTU) for their constant support and financial assistance. We are also thankful to the All India Institute of Medical Sciences (AIIMS) for SEM Analysis, the Department of Zoology, Delhi University for Gas Chromatography Analysis and the University Grants Commission (UGC) for providing fellowship support.

Author contribution All authors read and approved the final manuscript and also confirm the contribution to the research article: SM: investigation, methodology, data analysis and writing—original draft. RK: idea for the article, critically revised and editing the work and supervision. JGS: funding acquisition, resources and supervision.

Data availability Availability of data and materials could be provided if requested.

Declarations

Ethical approval Not applicable.

Competing interests The authors declare no competing interests.

References

1. Karunanithy C, Products KM (2011) Optimization of switchgrass and extruder parameters for enzymatic hydrolysis using response surface methodology. Ind Crops Prod 33(1):188–199
2. Sudhakar M, Ravel M, Perumal K (2021) Pretreatment and process optimization of bioethanol production from spent biomass of Ganoderma lucidum using Saccharomyces cerevisiae. Fuel 306:121680
3. Shrivastava A, Sharma RK (2023) Conversion of lignocellulosic biomass: production of bioethanol and bioelectricity using wheat straw hydrolysate in electrochemical bioreactor. Heliyon 9. https://doi.org/10.1016/j.heliyon.2023.e12951
4. Yadav KK, Krishnan S, Gupta N et al (2021) Review on evaluation of renewable bioenergy potential for sustainable development: bright future in energy practice in India. ACS Sustain Chem Eng 9:16007–16030. https://doi.org/10.1021/ACSSUSCHEMENG.1C03114
5. Kothari R, Vashishtha A, Singh H et al (2020) Assessment of Indian bioenergy policy for sustainable environment and its impact for rural India: strategic implementation and challenges. Environ Technol Innov 20:101078
6. Wyman CE, Yang B (2017) Combined severity factor for predicting sugar recovery in acid-catalyzed pretreatment followed by enzymatic hydrolysis. Hydrothermal Process Biorefineries Prod Bioethanol High Added-Value Compd Second Third Gener Biomass 161–180. https://doi.org/10.1007/978-3-319-56457-9_6
7. Kataria R, Woods T, Casey W et al (2018) Surfactant-mediated hydrothermal pretreatment of ryegrass followed by enzymatic saccharification for polyhydroxyalkanoate production. Ind Crops Prod 111:625–632
8. Olatunji KO, Ahmed NA, Ogunkunle O (2021) Optimization of biogas yield from lignocellulosic materials with different pretreatment methods: a review. Biotechnol Biofuels 14. https://doi.org/10.1186/S13068-021-02012-X
9. Sinbuathong N (2019) Predicting the increase of methane yield using alkali pretreatment for weeds prior to co-digestion. Energy Sources, Part A Recover Util Environ Eff 41:1124–1131. https://doi.org/10.1080/15567036.2018.1544990
10. Baadhe R, Potumarthi R, Technology NM (2014) Influence of dilute acid and alkali pretreatment on reducing sugar production from corncobs by crude enzymatic method: a comparative study. Bioresour Technol 162:213–217
11. Mikulski D, Kłosowski G (2023) Cellulose hydrolysis and bioethanol production from various types of lignocellulosic biomass after microwave-assisted hydrotropic pretreatment. Renew Energy 206:168–179. https://doi.org/10.1016/j.renene.2023.02.061
12. Phwan CK, Chew KW, Sebayang AH et al (2019) Effects of acids pre-treatment on the microbial fermentation process for bioethanol production from microalgae. Biotechnol Biofuels 12:1–8. https://doi.org/10.1186/S13068-019-1533-5/TABLES/2
13. Dahunsi SO (2019) Liquefaction of pineapple peel: pretreatment and process optimization. Energy 185:1017–1031. https://doi.org/10.1016/J.ENERGY.2019.07.123
14. Hendriks A, Technology GZ (2009) Pretreatments to enhance the digestibility of lignocellulosic biomass. Bioresour Technol 100(1):10–18
15. Gusain R, Suthar S (2017) Potential of aquatic weeds (Lemna gibba, Lemna minor, Pistia stratiotes and Eichhornia sp.) in biofuel production. Process Saf Environ Prot 109:233–241. https://doi.org/10.1016/J.PSEP.2017.03.030
16. John R, Anisha G, Nampoothiri K, Technology AP (2011) Micro and macroalgal biomass: a renewable source for bioethanol. Bioresour Technol 102(1):186–193
17. Kaur M, Kumar M, Singh D et al (2019) A sustainable biorefinery approach for efficient conversion of aquatic weeds into bioethanol and biomethane. Energy Convers Manag 187:133–147. https://doi.org/10.1016/j.enconman.2019.03.018
18. Whangchai K, Inta W, Unpaprom Y et al (2021) Comparative analysis of fresh and dry free-floating aquatic plant Pistia stratiotes via chemical pretreatment for second-generation (2G) bioethanol production. Bioresour Technol Reports 14:100651. https://doi.org/10.1016/J.BITEB.2021.100651
19. Myers RH, Montgomery DC, Anderson-Cook CM (2016) Response surface methodology: process and product optimization using designed experiments. Wiley
20. Montgomery DC (2017) Design and analysis of experiments. John wiley & sons
21. Mahalik K, Sahu JN, Patwardhan AV, Meikap B (2010) Statistical modelling and optimization of hydrolysis of urea to generate ammonia for flue gas conditioning. J Hazard Mater 182(1–3):603–610
22. Awoyale AA, Lokhat D (2021) Experimental determination of the effects of pretreatment on selected Nigerian lignocellulosic biomass in bioethanol production. Sci Rep 11:557. https://doi.org/10.1038/s41598-020-78105-8
23. Sluiter A, Hames B, Hyman D et al (2008) Determination of total solids in biomass and total dissolved solids in liquid process samples. Laboratory analytical procedure (LAP) (NREL)
24. Singh YD, Mahanta P, Bora U (2017) Comprehensive characterization of lignocellulosic biomass through proximate, ultimate and compositional analysis for bioenergy production. Renew Energy 103:490–500. https://doi.org/10.1016/J.RENENE.2016.11.039

25. Bauer S, Ibáñez AB (2014) Rapid determination of cellulose. Biotechnol Bioeng 111:2355–2357. https://doi.org/10.1002/BIT.25276
26. Kumar M, Turner S (2015) Protocol: A medium-throughput method for determination of cellulose content from single stem pieces of Arabidopsis thaliana. Plant Methods 11. https://doi.org/10.1186/S13007-015-0090-6
27. Wolfrum EJ, Lorenz AJ, deLeon N (2009) Correlating detergent fiber analysis and dietary fiber analysis data for corn stover collected by NIRS. Cellulose 16:577–585. https://doi.org/10.1007/S10570-009-9318-9/FIGURES/4
28. Holtzapple MT (2003) Hemicelluloses. In: Second E (ed) Caballero BBT-E of FS and N. Academic Press, Oxford, pp 3060–3071
29. Zhou Y, Zheng J, Gan R-Y et al (2017) Optimization of ultrasound-assisted extraction of antioxidants from the mung bean coat. Molecules. https://doi.org/10.3390/molecules22040638
30. Xia M, Valverde-Barrantes OJ, Suseela V et al (2022) Characterizing natural variability of lignin abundance and composition in fine roots across temperate trees: a comparison of analytical methods. New Phytol. https://doi.org/10.1111/NPH.18515
31. Nomanbhay SM, Hussain R, Palanisamy K et al (2013) Microwave-assisted alkaline pretreatment and microwave assisted enzymatic saccharification of oil palm empty fruit bunch fiber for enhanced fermentable sugar yield. J Sustain Bioenergy Syst 3:7–17. https://doi.org/10.4236/JSBS.2013.31002
32. Zorić AS, Morina F, Toševski I et al (2019) Resource allocation in response to herbivory and gall formation in Linaria vulgaris. Plant Physiol Biochem 135:224–232. https://doi.org/10.1016/J.PLAPHY.2018.11.032
33. Kataria R, Woods T, Casey W et al (2018) Surfactant-mediated hydrothermal pretreatment of ryegrass followed by enzymatic saccharification for polyhydroxyalkanoate production. Ind Crops Prod 111:625–632. https://doi.org/10.1016/J.INDCROP.2017.11.029
34. Umesh M, Santhosh AS, Shanmugam S et al (2022) Extraction, characterization, and fabrication of cellulose biopolymer sheets from Pistia stratiotes as a biodegradative coating material: an unique strategy for the conversion of invasive weeds into value-added products. J Polym Environ 30:5057–5068. https://doi.org/10.1007/S10924-022-02511-4/FIGURES/9
35. Gunaraj V, Technology NM (1999) Application of response surface methodology for predicting weld bead quality in submerged arc welding of pipes. J Mater Process Technol 88(1–3):266–275
36. Dahunsi SO, Adesulu-Dahunsi AT, Izebere JO (2019) Cleaner energy through liquefaction of cocoa (Theobroma cacao) pod husk: pretreatment and process optimization. J Clean Prod 226:578–588
37. Ahmed F, Yan Z, Bao J (2019) Dry biodetoxification of acid pretreated wheat straw for cellulosic ethanol fermentation. Bioresour Bioprocess. https://doi.org/10.1186/s40643-019-0260-x
38. Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem 31:426–428. https://doi.org/10.1021/AC60147A030
39. Bader NB, Germec M, Turhan I (2022) Repeated-batch fermentation of Scheffersomyces stipitis in biofilm reactor for ethanol production from the detoxified and glucose-or xylose-enriched rice husk hydrolysate and its kinetic modeling. Fuel 326:125053
40. Gonçalves FA, Ruiz HA, Silvino dos Santos E et al (2016) Bioethanol production by Saccharomyces cerevisiae, Pichia stipitis and Zymomonas mobilis from delignified coconut fibre mature and lignin extraction according to biorefinery concept. Renew Energy 94:353–365. https://doi.org/10.1016/J.RENENE.2016.03.045
41. Goswami RK, Sharma J, Shrivastav AK, Kumar G, Glencross BD, Tocher DR, Chakrabarti R (2022) Effect of Lemna minor supplemented diets on growth, digestive physiology and expression of fatty acids biosynthesis genes of Cyprinus carpio. Sci Rep 12(1):3711
42. Pereira SC, Maehara L, Machado CMM, Farinas CS (2015) 2G ethanol from the whole sugarcane lignocellulosic biomass. Biotechnol Biofuels 8. https://doi.org/10.1186/S13068-015-0224-0
43. Pooja NS, Sajeev MS, Jeeva ML, Padmaja G (2018) Bioethanol production from microwave-assisted acid or alkali-pretreated agricultural residues of cassava using separate hydrolysis and fermentation (SHF). 3 Biotech 8. https://doi.org/10.1007/S13205-018-1095-4
44. Mithra MG, Jeeva ML, Sajeev MS, Padmaja G (2018) Comparison of ethanol yield from pretreated lignocellulo-starch biomass under fed-batch SHF or SSF modes. Heliyon 4:885. https://doi.org/10.1016/J.HELIYON.2018.E00885
45. Sutaryo S, Sempana AN, Mulya RM et al (2022) Methane production of Pistia stratiotes as a Single substrate and as a co-substrate with dairy cow manure. Ferment 8:736. https://doi.org/10.3390/FERMENTATION8120736
46. Mishima D, Kuniki M, Sei K et al (2008) Ethanol production from candidate energy crops: water hyacinth (Eichhornia crassipes) and water lettuce (Pistia stratiotes L.). Bioresour Technol 99:2495–2500. https://doi.org/10.1016/J.BIORTECH.2007.04.056
47. Yosrey E, Elmansi H, Sheribah ZA, El-Sayed Metwally M (2021) Factorial design-assisted spectroscopic determination of oxybutynin hydrochloride. R Soc Open Sci 8. https://doi.org/10.1098/RSOS.211027
48. Auwal SM, Zarei M, Tan CP et al (2018) (2018) Enhanced physicochemical stability and efficacy of angiotensin I-converting enzyme (ACE) - inhibitory biopeptides by chitosan nanoparticles optimized using Box-Behnken design. Sci Rep 81(8):1–11. https://doi.org/10.1038/s41598-018-28659-5
49. Kataria R, Mol A, Schulten E et al (2017) Bench scale steam explosion pretreatment of acid impregnated elephant grass biomass and its impacts on biomass composition, structure and hydrolysis. Ind Crops Prod 106:48–58. https://doi.org/10.1016/j.indcrop.2016.08.050
50. Lu P, Hsieh YL (2010) Preparation and properties of cellulose nanocrystals: rods, spheres, and network. Carbohydr Polym 82(2):329–336
51. Luzi F, Puglia D, Sarasini F et al (2019) Valorization and extraction of cellulose nanocrystals from North African grass: Ampelodesmos mauritanicus (Diss). Carbohydr Polym 209:328–337
52. Mandal A, Chakrabarty D (2011) Isolation of nanocellulose from waste sugarcane bagasse (SCB) and its characterization. Carbohydr Polym 86(3):1291–1299
53. Hsu T, Guo G, Chen W, Hwang W (2010) Effect of dilute acid pretreatment of rice straw on structural properties and enzymatic hydrolysis. Bioresour Technol 101(13):4907–4913
54. Jmel M, Anders N, Messaoud G et al (2019) The stranded macroalga Ulva lactuca as a new alternative source of cellulose: extraction, physicochemical and rheological characterization. J Clean Prod 234:1421–1427
55. Alemdar A, Sain M (2008) Isolation and characterization of nanofibers from agricultural residues–wheat straw and soy hulls. Bioresour Technol 99(6):1664–1671
56. Deng Z, Xia A, Liao Q et al (2019) Laccase pretreatment of wheat straw: effects of the physicochemical characteristics and the kinetics of enzymatic hydrolysis. Biotechnol Biofuels 12. https://doi.org/10.1186/S13068-019-1499-3
57. Malik K, Salama E, Kim T, Li X (2020) Enhanced ethanol production by Saccharomyces cerevisiae fermentation post acidic and alkali chemical pretreatments of cotton stalk lignocellulose. Int Biodeterior Biodegrad 147:104869
58. Kale RD, Taye M, Chaudhary B (2019) Extraction and characterization of cellulose single fiber from native Ethiopian Serte (Dracaena steudneri Egler) plant leaf. J Macromol Sci Part A Pure

Appl Chem 56:837–844. https://doi.org/10.1080/10601325.2019.1612252
59. Sunwoo IY, Kwon JE, Nguyen TH et al (2019) Ethanol production from water hyacinth (Eichhornia crassipes) hydrolysate by hyperthermal acid hydrolysis, enzymatic saccharification and yeasts adapted to high concentration of xylose. Bioprocess Biosyst Eng 42:1367–1374. https://doi.org/10.1007/S00449-019-02136-3
60. Sulaiman NMA, El-Mohsen A, Refaat A et al Kinetics of bioethanol production on the molasses-based medium by Saccharomyces cerevisiae. https://doi.org/10.21608/asajs.2022.228846
61. Tran PHN, Jung JH, Ko JK et al (2023) Co-production of ethanol and polyhydroxybutyrate from lignocellulosic biomass using an engineered Saccharomyces cerevisiae. Renew Energy 212:601–611. https://doi.org/10.1016/J.RENENE.2023.05.080
62. Nunui K, Boonsawang P, Chaiprapat S, Charnnok B (2022) Using organosolv pretreatment with acid wastewater for enhanced fermentable sugar and ethanol production from rubberwood waste. Renew Energy 198:723–732. https://doi.org/10.1016/J.RENENE.2022.08.068
63. Irfan M, Nadeem M, Syed Q (2014) Ethanol production from agricultural wastes using Sacchromyces cervisae. Brazilian J Microbiol 45:457. https://doi.org/10.1590/S1517-83822014000200012
64. Kataria R, Ghosh S (2011) Saccharification of Kans grass using enzyme mixture from Trichoderma reesei for bioethanol production. Bioresour Technol 102(21):9970–9975

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 | FEBRUARY 20 2024

Performance parameter extraction of SnZnTe absorber layer photovoltaic cell FREE

Paritosh Chamola ✉; Poornima Mittal

Check for updates

AIP Conf. Proc. 2986, 030125 (2024)
https://doi.org/10.1063/5.0192572

View Online Export Citation CrossMark

26 February 2024 07:51:00



AIP Publishing

Performance Parameter Extraction of SnZnTe Absorber Layer Photovoltaic Cell

Paritosh Chamola[1, a)] and Poornima Mittal[1, b)]

[1] *Department of Electronics and Communication Engineering, Delhi Technological University, Delhi-110042 (India)*

[a)] Corresponding author: paritoshchamola@gmail.com
[b)] poornimamittal@dtu.ac.in

Abstract. The solar cells are the eco-friendly, renewable and best alternate available to the presently available non-renewable and pollution creating form of energy. The photovoltaic cells capture solar energy or photons and convert it to electrical current. The thin-film photovoltaic devices based on semiconductors are low cost and effective devices for generating current from the solar energy. The energy of photons depends on the light's wavelength, in a semiconductor material electric current is generated when an incoming photon is able to incite an electron present at the valance to conduction band. In this article ZnTe (Zinc Telluride) heterojunction photovoltaic cell is presented and its performance parameters are obtained and further parameter optimization is performed. In the present photovoltaic cell SnZnTe material work as absorber layer. ZnO (Zinc Oxide) material is window layer and CdS (Cadmium-Sulphide) act as buffer layer. The simulation studies are performed via Silvaco Atlas 2D device simulator and different important device parameters such as fill factor (FF), efficiency (η), open circuit voltage (V_{OC}) etc. are obtained followed by device optimization. The Silvaco Atlas 2D device simulator is able to present an extremely detailed internal device physics. Photovoltaic cell's characteristic properties can be easily observed by the Silvaco device simulator. In present article SnZnTe photovoltaic cell is studied and its simulation is performed the SnZnTe photovoltaic cell shows commendable performance with a fill factor of 75.2%.

Keywords - Photovoltaic cell, Fill Factor, Open circuit voltage, semiconductor, Power conversion efficiency.

26 February 2024 07:51:00

INTRODUCTION

The simulation tools are of vital importance, all the important and critical aspects of photovoltaic device can be understood by using these simulation model software's. Further, all the optimization needed for any type of improvements in the solar cell can be performed by using mathematical simulators [1]. Elements in second and sixth group turn out to be excellent in manufacturing of inexpensive solar cell devices for example ZnTe (Zinc Telluride) showing a band-gap of 2.24 eV [2], electron affinity of 3.73 eV [3]. In general, when Zinc Telluride is doped with first and fifth group elements or with rare materials such as Sn (Tin) which leads to better properties of the photovoltaic cell. Thus, performance of the ZnTe photovoltaic cell can be improved if doped with Sn (Tin) material to form SnZnTe solar cell.

The photovoltaic cells of the first generation are based on simple wafers and are of single crystalline nature. The second generation of photovoltaic device is represented by depositing the semiconducting substance on a glass substrate. The photons having energy lesser than the band gap of absorber material are left unabsorbed and hence do not produce current. The third generation of photovoltaic cell are presented by solar cell having intermediate band. This intermediate band is formed by alloying the mismatched elements and host element [4]. In present article SnZnTe photovoltaic cell is studied and its simulation is performed and performance parameters are extracted using Silvaco Atlas 2D device simulator.

This paper is arranged in four sections with the introduction as first section. The second section presents photovoltaic cells structure and materials used in layer structure of the cell. In the section third parameter extraction

3rd International Conference on Functional Materials, Manufacturing, and Performances
AIP Conf. Proc. 2986, 030125-1–030125-6; https://doi.org/10.1063/5.0192572
Published by AIP Publishing. 978-0-7354-4840-7/$30.00

and their optimization for best performing photovoltaic cell is accomplished [5]. Finally, the section four presents all the important results.

DEVICE SIMULATION METHODOLOGY

The Silvaco Atlas simulator allows user to define their own materials and various types of models are available for simulating the solar cells. The photovoltaic cell presented in this article uses Poole-Frenkel for mobility and Shockley-Read-Hall (SRH) model calculations for the simulation, due to the availability of different types of optical propagation models [6] available in the Silvaco software hence the solar cells simulated are of high accuracy.

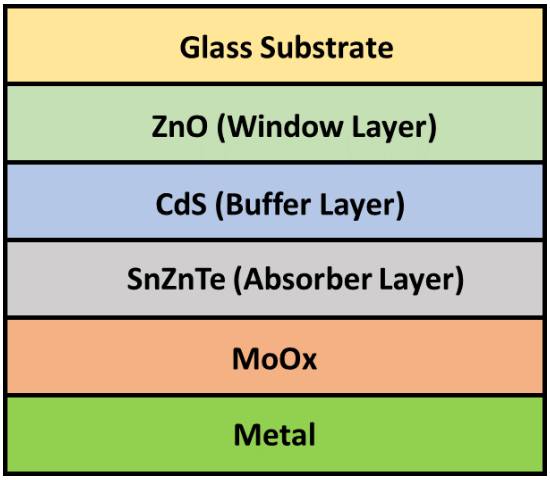


FIGURE 1. SnZnTe photovoltaic cell.

The Figure. 1 illustrates the structure of simulated photovoltaic cell. The structural design and the purpose of the photovoltaic cell is the most prominent aspects as per which the optical models are decided to simulate the cell and the complete simulation results are unconditionally reliant on these models. The alignment of energy bands is the most principal aspect that can heavily imapct the carrier transport of generated charge carriers and photovoltaic cell's functionality. The energy band schema are shown up in Fig. 2 for SnZnTe photovoltaic cell illustrating a good band alignment.

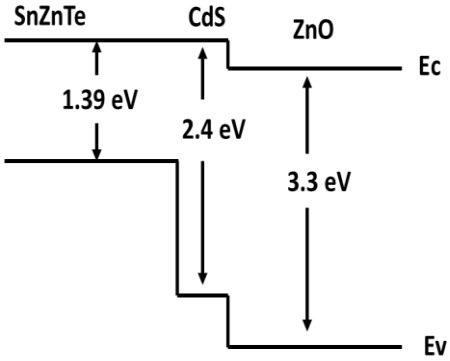


FIGURE 2. Energy band diagram for SnZnTe photovoltaic cell.

26 February 2024 07:51:00

Conditional on the current-voltage characteristics the photovoltaic parameters namely efficiency(η), short circuit current (J_{SC}), open circuit voltage (V_{OC}), fill factor (FF) and maximum power density (P_m) can be calculated.

$$FF = Pm / V_{OC}.J_{SC} \quad (1)$$

$$\text{Where, } P_m = V_m . J_m \quad (2)$$

$$\eta = (P_m \text{ X } 100\%) / \text{Irradiance} \quad (3)$$

TABLE 1. Material parameters and layer thickness for SnZnTe photovoltaic cell

Material parameters (unit)	**Default Values**		
	SnZnTe	CdS	ZnO
Band-gap (eV)	1.39	2.4	3.3
Thickness (μm)	1.5	0.060	0.050
Permittivity (ε_r)	10	10	9
Electron mobility ($cm^2/V. s$)	340	350	100
Hole mobility ($cm^2/V. s$)	100	40	25
Electron affinity (eV)	3.50	4.5	4.6

All the simulation parameters are summarized in table 1.

RESULTS AND DISCUSSION

The material of absorber layer has an extremely important role to play in defining and determining the photovoltaic cells performance. Thus, Sn material is doped to the ZnTe forming SnZnTe material and this SnZnTe is used in solar cell to obtain the improved performance characteristics. Thus, in the present photovoltaic cell SnZnTe is used as a material for the absorber layer and performance of SnZnTe photovoltaic cell is observed as well as analyzed by keeping the other layers of the photovoltaic cell constant.

The Fig. 3 presents the simulated solar cell designed in Silvaco Atlas simulation software the different layers are shown via different colors (i) ZnO is presented by yellow color (ii) CdS by green color (iii) SnZnTe via pink color layer and (iv) blue color represents the conductors

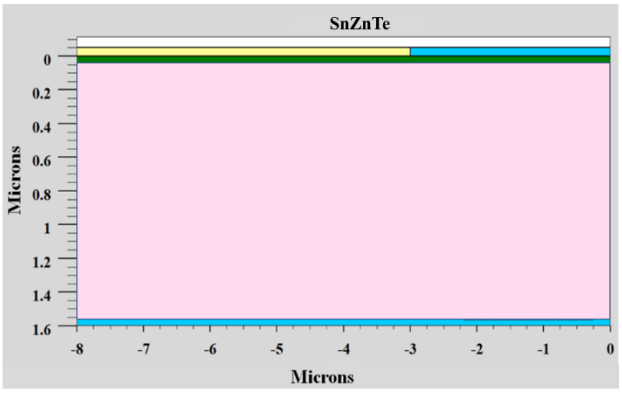


FIGURE 3. Device structure of the simulated SnZnTe solar cell

26 February 2024 07:51:00

The performance parameters extracted by simulating the SnZnTe absorbing material based photovoltaic device. The Fig. 4 represents the current density analyzed at different voltages for the SnZnTe photovoltaic cell. The current polarity is analyzed to be positive and is similar to that of described by various researchers such as Khattak *et al.* [4], Tanaka *et al.* [7, 8], Dharmadasa *et al.* [9] and Mukhopadhyay et al. [10].

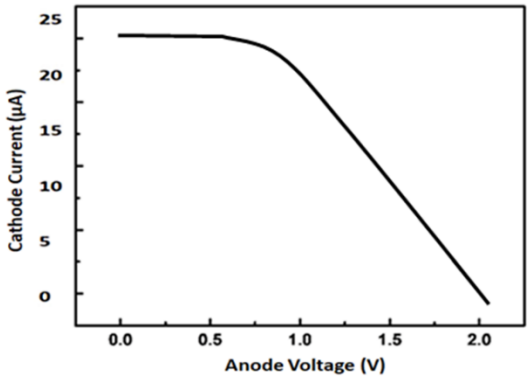


FIGURE 4. Current-voltage curve for SnZnTe solar cell.

The peak J_{SC} of 23.3 $mAcm^{-2}$ is noticed in SnZnTe cell at a voltage of 0 volts, when device is short circuit i.e., no load is connected to the circuit. This significant improvement in current is owed to SnZnTe material's small bandgap of 1.39 eV, which is suitable for material substances used to make photovoltaic cell.

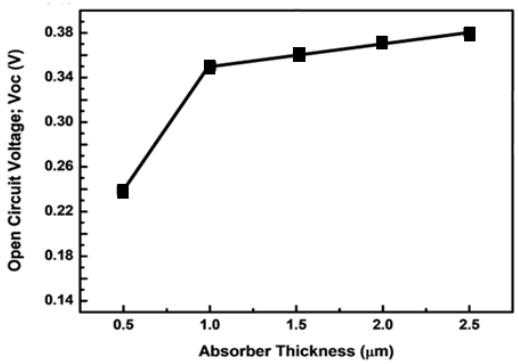


FIGURE 5. The characteristic plot of Open circuit voltage versus absorber thickness in an SnZnTe photovoltaic cell.

Figure. 5 illustrate the open circuit voltage plots at different thickness of SnZnTe photovoltaic cell respectively. The trend shown by the open-circuit voltage can be described as linear rise, V_{OC} rises from 0.24 to 0.42 V when absorber layer's thickness is increased (0.5 - 2.5 μm). The V_{OC} gradual slope remains so owing to the restricted increase in number of recombination sites. Prominent factors determining the open circuit voltages are fermi splitting and the rate of carrier recombination.

26 February 2024 07:51:00

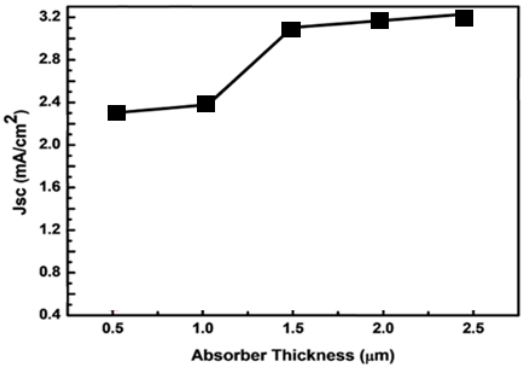


FIGURE 6. Characteristic plot of short-circuit current density versus absorber thickness for SnZnTe photovoltaic cell.

The J_{SC} in a photovoltaic device is calculated at 0 Volt. The short circuit parameter of a photovoltaic cell are dependent on various aspects for instance incident power of photons, cell's area, incident light's spectrum, reflection and absorption optical properties of the cell. In the fig. the short circuit current increases with the increasing absorber thickness due to more absorption of light and thus a greater number of free charge carriers are generated, this remains prominent till absorber thickness of 1.5 μm as maximum charge carriers are adjacent to interface neighbourhood. The current density is also limited by charge carrier recombination. The maximum value of 3.2 mA/cm^2 is obtained at a thickness of 1.5μm.

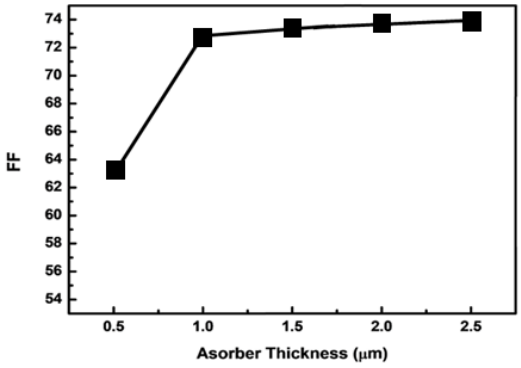


FIGURE. 7. Characteristic plot of Fill Factor versus absorber thickness for SnZnTe photovoltaic cell.

The fill factor (FF) is the utmost crucial performance parameter of a photovoltaic cell as it presents the quality of the photovoltaic cell and also determines the photovoltaic cell's efficiency. Therefore, it's becomes the prominent performance parameter in assessing the solar cell's performance. The FF versus absorber layer thickness is presented in fig 6. The attained value for SnZnTe is 75.2% at a thickness of 1.5 μm, from the plot in fig.6 it can be seen that FF demonstrates a little surge while increasing the thickness of the absorber. This increase is due to the growth in the number of charge carriers produced due to generation of higher number of electron hole pair.

CONCLUSION

The present article emphasizes on detailed examination of Sn doped ZnTe photovoltaic cell. The SnZnTe photovoltaic device is simulated as well as analyzed to optimize device performance. The all-performance parameters for instance open circuit voltage, short circuit current, and fill factors are extracted. The results obtained from the simulations present the optimal absorber thickness for outstanding operation of photovoltaic device. A fill factor of 75.2 % is obtained showing that SnZnTe suits best as an absorber layer material also from the variation of

26 February 2024 07:51:00

absorber layer thickness it can be shown that the absorber layer thickness is extremely influential in refining the performance of photovoltaic cell without changing the device dimensions.

REFERENCES

1. N. Amin, K. Sopian and M. Konagai, Sol. Energy Mater. and Sol. Cell, **91**, 1202-1208 (2007).
2. A. Kaneta and S. Adachi, J. Phys. D: Appl. Phys. **33**, 901-905 (2000).
3. A. Pistone, A. S. Arico, P. L. Antonucci, D. Silvestro and V. Antonucci, Sol. Energy Mater. Sol. Cell, **53**, 255-267(1998).
4. Y. H. Khattak, F. Baig, S. Ullah, B. Marí, S. Beg and H. Ullah, J. of Renew. and Sustain. Energy, **10**, 033501-1-033501-14 (2018).
5. ATLAS (2014) ATLAS User's Manual Device Simulation Software, Silaco International Ltd, Santa Clara.
6. P. Mittal, Y. S. Negi and R. K. Singh, J. of Com. Elec., **14**, 360-379 (2015).
7. T. Tanaka, M. Miyabara, K. Saito, Q. Guo, M. Nishio, K.M. Yu and W. Walukiewicz, Mater. Sci. Forum, **750**, 80–83 (2013).
8. T. Tanaka, K.M. Yu, P.R. Stone, J.W. Beeman and O.D. Dubon, J. Appl. Phys. **108,** 024502-024502-3 (2010).
9. I.M. Dharmadasa, Y. Rahaq, A.A. Ojo, T.I. Alanazi, J Mater Sci: Mater. Electron. **30**, 1227–1235 (2019).
10. K. Mukhopadhyay, P.F.H. Inbaraj and J.J. Prince, Mater. Res. Innov. **23**, 1–11 (2018).

26 February 2024 07:51:00

Indian Journal of Traditional Knowledge
Vol 23(2), February 2024, pp 162-169
DOI: 10.56042/ijtk.v23i2.3575

Phytochemical, antioxidant and antimicrobial activity of *Livistona chinensis* (Jacq.) R.Br. ex Mart. seeds

Tamanna Beniwal, Ritu Sharma & Rajinder K Gupta*
Department of Applied Chemistry, Delhi Technological University, Delhi 110 042, India
*E-mail: rkg67ap@yahoo.com

Received 01 August 2023; revised 28 December 2023; accepted 30 January 2024

The seeds of *Livistona chinensis* (Jacq.) R.Br. ex Mart. (Family: Arecaceae) plant traditionally find application in the treatment of various diseases. Therefore, in the present study, phytochemical screening and biological profiling of *Livistona chinensis* seed extracts were conducted. The bioactive components were analysed using GC-MS and UHPLC-QTOF-MS and the fatty acid composition was determined using FAME analysis. GC-MS analysis results indicated the presence of oleic acid (35.71%) and palmitic acid (30.7%) as major components in the seed extract prepared in n-hexane. UHPLC-QTOF-MS analysis revealed the presence of 25 secondary metabolites, including isobutyric acid, acetoin, 9-oxononanoic acid, piperidine, umbellic acid, furaneol, and ketopantolactone in the methanolic extract of seeds. FAME analysis results suggested the presence of oleic acid (18.44%), lauric acid (11.66%), linoleic acid (11.30%), linolenic acid (10.20%), and palmitic acid (9.96%) in abundance. In biological profiling, *Livistona chinensis* seed extracts were subjected to *in vitro* antimicrobial, and antioxidant activity. The aqueous extract exclusively showed antimicrobial activity, and exhibited inhibition zone of 16.1 mm only against gram-negative bacteria (*Klebsiella pneumonia*). The methanolic extract demonstrated maximum antioxidant activity, at sample concentration of 200 µL. These findings suggest that *Livistona chinensis* seeds might be a potential source of bioactive metabolites, antimicrobial agents, and antioxidant agents that may be utilised in pharmaceutical, cosmetics, and food industries.

Keywords: Antimicrobial activity, Antioxidant activity, Elemental analysis, *Livistona chinensis*, Phytochemical analysis, UHPLC-QTOF-MS

IPC Code: Int Cl.[24]: A61K 36/00

Nutraceutical and functional foods are essential in managing numerous lifestyle-related illnesses like diabetes, obesity, and heart disease. Due to the abundance of health-promoting ingredients present today, goods from plant sources are attracting a lot of interest. A huge attention has shifted towards medicinal seeds in particular due to their rich phytochemistry and preventive properties[1].

Livistona chinensis (*L. chinensis*, Family: Arecaceae) is a decorative plant and is recognised as Chinese Fountain palm or Chinese Fan palm. Its native habitats are Taiwan, Southern Japan, East Asia, Australia, and several islands in the Southern China Sea[2]. The fruits of *L. chinensis* are oval in shape, with a hard outer covering, and black in colour. According to the previous reports, *L. chinensis* fruits exhibited anti-cancer, anti-microbial, and anti-HIV-1 properties[3]. Historically, the analgesic and haemostatic properties of their fruits have been utilised to treat nasopharyngeal carcinoma, choriocarcinoma, oesophageal cancer, gastric cancer, and leukemia[4]. People in China and East Asia have utilised the *L. chinensis* fruit as an herbal ingredient in soups to treat hepatitis and many types of liver cancer[5]. *L. chinensis* seeds have a long history of usage in Traditional Chinese Medicine, as an anti-cancer agent. Also, various experimental studies have shown the anti-proliferative and anti-angiogenic properties of aqueous extract from their seeds and fruits[6].

Moreover, the presence of phenolic compounds in fruits of *L. chinensis* has been reported to possess haemolytic activity[7]. Chinese Fan palm seeds' biochar were also found to absorb the Malachite Green, which is an effluent from the textile dye[8].

Phytochemicals, mainly found in plants are secondary metabolites and are responsible for their different medicinal properties. The preliminary screening of phytochemicals was crucial in discovering the different phytoconstituents present in

*Corresponding author

medicinal plant extracts, providing an insight into future drug discovery and development[9]. The chemical constituents present in plant parts are considered as biologically active compounds as they possess diverse activities like anti-analgesic, anti-cancer, anti-viral, and anti-microbial[10]. The phytoconstituents like alkaloids, glycosides, saponins, resins, and oils have a protective or disease-preventive effect[11].

Medicinal plants are now gaining more credibility due to the growing demand for their antioxidant properties. In the food industry, anti-oxidants are used to delay the oxidation process. Natural substances are the possible substitute for the widely used synthetic anti-oxidants. In this paper, the anti-oxidant activity of the seeds extract was evaluated using free radical 2,2-diphenyl-1-picrylhydrazyl ($DPPH^{\cdot}$), which absorbs strongly at 517 nm, and its absorbance decreases with the addition of any antioxidant[12].

Fatty acid composition controls functional and nutritional values. Polyunsaturated fatty acids (PUFA) like linoleic (omega-6) and linolenic acids (omega-3) are crucial for human metabolism due to a shortage of the enzymes needed for their formation. PUFA are regarded as desirable substances in the human diet due to their ability to lower the risk of cancer and cardiovascular disease[13]. The novelty in this study is the application of soxhlet extraction of seeds of *L. chinensis* followed by the use of ultra-high performance liquid chromatography-quadrupole time-of-flight mass spectrometry (UHPLC-QTOF-MS) method to analyse the secondary metabolites present in the isolated extract.

In view of the above, the current study investigated the evaluation of the bioactive compounds of *L. chinensis* seeds cultivated in India by UHPLC-QTOF-MS and GC-MS analysis, elemental analysis by ICP-MS analysis, antimicrobial property against *Klebsiella pneumoniae* (gram-negative) and *Staphylococcus aureus* (gram-positive), and antioxidant property by DPPH free radical scavenging method.

Material and Methods

Plant material collection and authentication

Livistona chinensis fruits were collected from the trees growing at Delhi Technological University, New Delhi, India. The botanical specimens of the *Livistona chinensis* fruits were authenticated by CSIR-NIScPR, Raw Materials Herbarium, and Museum, Delhi (RHMD), India and the voucher specimen (NIScPR/RHMD/Consult/2023/4482-83) was deposited at the RHMD, Delhi. The collected *L. chinensis* fruits were cleaned and their seeds were removed manually from the fruits. The seeds were sun dried for 2-3 days and coarsely grounded by grinder for their extraction.

Preparation of extracts

L. chinensis (50 g) grounded seeds were extracted sequentially with n-hexane, dichloromethane (DCM), methanol, and deionised water for 6-8 h, using Soxhlet apparatus. After successful extraction, the solvents were evaporated using a rotary evaporator. The concentrated extracts obtained were kept at 4°C for further use[14].

Phytochemical analysis

The qualitative phytochemical analysis of methanol extract of *L. chinensis* seeds was done following the standard methods[15-17].Various phytoconstituents identification tests for tannins, alkaloids, glycosides, anthraquinones, steroids, proteins, flavonoids, terpenoids, reducing sugars, anthocyanins, emodins, coumarins, phlobatannins, and phenols were performed.

GC-MS (Gas Chromatography-Mass Spectrometry) analysis

The determination of volatile compounds present in the crude n-hexane extract of *L. chinensis* seeds was carried out using GC-MS (Shimadzu, GCMS-QP2010). The sample was injected into GC-MS, using Helium as a carrier gas in split mode at 260°C. The column flow rate was kept at 1.21 mL/min and the ionization mode was electron ionisation (EI). The ion source and interface temperature were 220°C and 270°C, respectively. A solvent delay time of 3.50 min, the temperature of the oven was set at 80°C for 2 min to 300°C for 21 min. To identify the separated peaks, NIST mass spectral data-base (National Institute of Standards and Technology) was used[18].

UHPLC-QTOF-MS analysis

For the UHPLC and mass spectrometric analysis of the methanolic extract of *L. chinensis* seeds, HPLC (Waters, SYNAPT-XS HDMS, UK) fitted with an AD pump, AD auto sampler controller, degasser, AD column, coupled with a quadrupole time-of-flight mass spectrometer (QTOF-MS) was used. It was used for the identification of secondary metabolites in the DCM and methanol extracts of the seeds. Both the extracts were mixed with formic acid (1%, 10 mL) in water and set aside for 10 min. 10 mL acetonitrile and

10 mL methanol were mixed and vortexed for 1 min then centrifuged at 5000 rpm for 5 min. Diluting the sample with acidified water, it was injected (5 µL) into the instrument. 100 mm x 2.1 mm column C18 (Waters, Acquity BEH 2.1) was used for the chromatographic separation of the samples. The mass spectroscopic analysis was done in only positive mode (ESI+) with the following mentioned conditions: desolvation gas flow (950 L/h), source temperature, (120°C), capillary voltage (3.22 keV), cone gas flow (50 L/h). For eluting the secondary metabolites, a binary mobile phase at a flow rate of 0.2 mL/min was used, where solvent A was LC-MS grade water containing 1% formic acid and solvent B was 1% formic acid with acetonitrile. ChemSpider software was used to perform data acquisition and processing[19].

ICP-MS (Inductively Coupled Plasma Mass Spectrometry) analysis

About 0.1 g of the *L. chinensis* grounded seeds were taken in digestion vessel. 1 mL H_2O_2, 1 mL distilled water, and 2 mL H_2SO_4 (conc.) was added. After 30 min, the digestion vessel was closed and kept within the digester at 483 K. The similar method was used to examine blank sample and NIST standards. The instrument calibration was evaluated using three NIST-certified samples[20].

FAME (Fatty Acid Methyl Ester) analysis

The fatty acid content in the *L. chinensis* seeds were evaluated using Fatty acid methyl ester (FAME) analysis by taking 0.2 mL of the extracted oil sample. 200 µL of 2N alcoholic potassium hydroxide (KOH) was used to saponify the oil obtained from the seeds and vortexed for 2 min. This mixture was heated for 10 min in a water bath at 60°C, and then cooled off. To obtain mixed fatty acids (MFAs), the aqueous layer was then acidified with 1 mL of 5% hydrochloric acid (HCl) solution, heated for 10 min at 70°C. The mixture obtained was cooled and vortexed for 2 min after addition of 2 mL of petroleum ether. The separated upper layer was examined by GC-MS (Shimadzu, GC-2010).

Antioxidant activity

DPPH radical scavenging activity

The method used for antioxidant activity described by Velazquez *et al*., was applied with some modifications[13]. The methanolic extract of the sample at different concentrations (60 µL, 80 µL, 100 µL, 120 µL, 140 µL, 160 µL, 180 µL, and 200 µL) were made and 1 mL of each sample was mixed with 3 mL of DPPH radical, with methanol serving as the blank sample. The samples were incubated for 30 min at 37°C, and after that the absorbance was measured for each sample at 517 nm. The % radical scavenging was measured using the following formula[12]:

$$\% \text{ Scavenging activity} = \frac{\text{Absorbance of Control} - \text{Absorbance of sample}}{\text{Absorbance of Control}} \text{ x } 100$$

Antimicrobial activity

The antimicrobial activity of aqueous and methanol extracts of *L. chinensis* were done using the disc diffusion method[21]. 20 mL of the nutrient media was poured into sterilised petri dishes, which was then kept to solidify. 20 µL of the extract was tested against following bacterial strain: gram-negative bacteria, *Klebsiella pneumoniae* and gram-positive bacteria, *Staphylococcus aureus*. The bacterial strains were cultivated at 37°C in incubator and left for 24 h. After the incubation period, zone of inhibition was measured.

Results and Discussion

Phytochemical analysis of *L. chinensis* seeds extract

The data of phytochemical analysis revealed that tannins, glycosides, alkaloids, phenols, and emodins were present in water and methanol extract, while flavonoids, terpenoids, reducing sugars, steroids, anthraquinones, and coumarins were only identified in the methanol extract (Table 1). Proteins, anthocyanins, and phlobatannins were not identified in any of the extracts. The phytoconstituents present in both the extracts were identified as biologically active compounds and they have been responsible for

Table 1 — Phytochemical analysis of *L. chinensis* seeds extracts

S. No.	Phytochemicals	Result	
		Methanol Extract	Water Extract
1.	Tannins	++	++
2.	Flavonoids	+	-
3.	Glycosides	++	+
4.	Terpenoids	+	-
5.	Reducing Sugar	+	-
6.	Proteins	-	-
7.	Anthocyanins	-	-
8.	Steroids	+	-
9.	Anthraquinones	+	-
10.	Emodins	+	+
11.	Alkaloids	+++	+
12.	Phlobatannins	-	-
13.	Coumarins	+	-
14.	Phenols	++	++

+++ = strongly presence of phytochemicals, + = presence of phytochemicals, - = absence of phytochemicals

various activities. Phenolic compounds are well-known for antiviral, antioxidant, anti-mutagenic, antimicrobial activities. Terpenoids also show anti-bacterial, anti-malarial, anti-inflammatory, antiviral, anti-cancer, and antimicrobial activities[10]. Hifnawy *et al*.[22], analysed similar results for the phytochemical screening of seeds extract of *L. chinensis*.

Gas chromatography-mass spectrometry analysis of *L. chinensis* seeds extract

Fifty-five volatile components were characterised in the n-hexane extract of *L. chinensis* seeds by GC-MS studies (Table 2). The major constituents were found to be oleic acid (35.71%) and palmitic acid (30.7%). These phytochemicals may be responsible for diverse pharmacological activities like hepatoprotective, antioxidant, wound-healing, and antibacterial activities.

UHPLC-QTOF-MS analysis

UHPLC-QTOF-MS analysis was done to evaluate secondary metabolites present in the DCM and methanol extract of *L. chinensis* seeds. UHPLC-QTOF-MS data of methanol and DCM extract as shown in

Table 2 — Bioactive compounds identified by GC-MS analysis of n-hexane extract of *L. chinensis* seeds

Name of Compounds	RT (min)	Area %	MW ($gmol^{-1}$)	MF
3-Trimethylsilyloxy-24-ethylcholesta-5,22-diene	27.437	3.41	484	$C_{32}H_{56}OSi$
3β-(Trimethylsiloxy) stigmasta-5,22-diene	26.748	0.34	484	$C_{32}H_{56}OSi$
1-(Trimethylsilyloxy) octacosane	25.343	0.97	482	$C_{31}H_{66}OSi$
2-(Dodecanoyloxy)-1- ([(trimethylsilyl) oxy] methyl) ethyl laurate	25.11	0.57	528	$C_{32}H_{60}O_5Si$
2-Tetradecyloxirane	24.791	0.14	240	$C_{16}H_{32}O$
Pentatriacontane	24.298	0.41	450	$C_{32}H_{66}$
Squalene	23.039	0.62	410	$C_{30}H_{50}$
Carbonic acid, eicosyl prop-1-en-2-yl ester	22.923	0.35	382	$C_{24}H_{46}O_3$
Octacosane	21.462	0.11	394	$C_{28}H_{58}$
Dioctyl phthalate	20.993	0.27	390	$C_{24}H_{38}O_4$
2-Methyleicosane	20.69	0.21	296	$C_{21}H_{44}$
(3Z,13Z)-2-Methyl-3,13-octadecadien-1-ol	20.545	0.24	280	$C_{19}H_{36}O$
Tricosane	19.89	0.09	324	$C_{23}H_{48}$
2-Butyloctanol	19.059	0.38	186	$C_{12}H_{26}O$
Tributylacetylcitrate	18.604	0.08	402	$C_{20}H_{34}O_8$
Trimethylsilyl stearate	18.531	1.01	356	$C_{21}H_{44}O_2Si$
Oleic acid, trimethylsilyl ester	18.342	35.71	354	$C_{21}H_{42}O_2Si$
Octadecyl trimethylsilyl ether	17.753	0.36	342	$C_{21}H_{46}OSi$
Heptadecanoic acid, trimethylsilyl ester	17.654	0.16	342	$C_{20}H_{42}O_2Si$
Methyl linolelaidate	17.253	0.33	294	$C_{19}H_{34}O_2$
Palmitic acid, trimethylsilyl ester	16.766	30.7	328	$C_{19}H_{40}O_2Si$
Palmitoleic acid 1tms	16.536	0.17	326	$C_{19}H_{38}O_2Si$
Octadecane	16.333	0.17	254	$C_{13}H_{28}$
Ethyl undecanoate	16.277	0.04	214	$C_{13}H_{26}O_2$
1-Trimethylsiloxyhexadecane	15.900	0.18	314	$C_{19}H_{42}OSi$
Trimethylsilyl pentadecanoate	15.783	0.15	314	$C_{18}H_{38}O_2Si$
Methyl hexadecanoate	15.621	0.09	270	$C_{17}H_{34}O_2$
17-Octadecynoic acid, trimethylsilyl ester	15.496	0.10	352	$C_{21}H_{40}O_2Si$
n-Decyl fluoride	15.29	0.42	160	$C_{10}H_{21}F$
Trimethylsilyl myristate	14.782	2.79	300	$C_{17}H_{36}O_2Si$
Isopropyl tetradecanoate	14.559	0.35	270	$C_{17}H_{34}O_2$
Hexadecane	14.308	0.33	226	$C_{16}H_{34}$
Trimethylsilyl laurate	12.64	3.55	272	$C_{15}H_{32}O_2Si$
Heptadecane	12.069	0.55	240	$C_{17}H_{36}$
Tridecane	9.563	0.68	184	$C_{13}H_{28}$
Dodecane	6.746	0.29	170	$C_{12}H_{26}$
Trimethylenenorbornane	5.347	0.11	136	$C_{10}H_{16}$
Di hydrodicyclopentadiene	4.954	0.13	134	$C_{10}H_{14}$

RT = Retention time
MF = Molecular formula

Table 3 confirms the presence of chloroneb, isobutyric acid, acetoin, 9-oxononanoic acid, piperidine, umbellic acid, L-(+)-mandelic acid, 2-hydroxyphenylacetic acid, vanillin, 3-hydroxyphenyl-acetic acid, 3,4-dihydroxyphenylacetaldehyde, artemisinic aldehyde, zerumbone, solavetivone, 4-hydroxyphenylpyruvate, trans-caffeic acid, 4-hydroxyphenylacetic acid, butyl propionate, (9E)-9-octadecenedioate, 1-naphthyl β-D-glucopyranoside,1,2-dihydropyrimidine, 5',5'-dimethyl-1,1'-bi(cyclohexane)-1',3-diene-4-carbaldehyde, 2-(2-methylenecyclopropyl)-3-oxo-succinic acid, furaneol, and ketopantolactone.

According to the reported study, isobutyric acid plays an important role in maintaining improving intestinal barrier function, reducing risk of chronic diseases (inflammatory bowel disease, colorectal cancer, and type 2 diabetes) because of its anti-inflammatory and antioxidant properties. It also enhances the activity of key enzymes involved in the synthesis of tight junction proteins[23]. Acetoin shows antioxidant and anti-inflammatory properties useful for maintaining cellular health and preventing disease. It is an intermediate in the biosynthesis of several amino acids, including valine, leucine, and isoleucine. It is used as a flavouring agent in food industry and used for fragrance and fixative in the perfume industry[24]. 9-oxononanoic acid has antimicrobial and anticancer properties. Piperidine acts as a neurotransmitter, and exhibits analgesic, antibacterial, antifungal, anti-inflammatory, and antitumour activities having wide applications in the field of pharmaceuticals and agrochemicals[25].

ICP-MS analysis of *L. chinensis* seeds

The elemental analysis of *L. chinensis* seeds, as presented in Table 4, showed the presence of different types of minerals at different concentrations The main elements in *L. chinensis* seeds are iron, zinc and calcium. Calcium was the most abundant macro-

Table 3 — UHPLC-QTOF-MS of DCM and methanol extract of *Livistona chinensis* seeds

Peak No.	Tentative metabolites	RT (min)	MF	MM ($gmol^{-1}$)	$[M-H]^+$	Error (ppm)	Compound ID
DCM extract							
1.	Chloroneb	9.389	$C_8H_8Cl_2O_2$	205.989	206.997	0.085	CSID16623
2.	Isobutyric acid	10.38	$C_4H_8O_2$	88.0522	89.06	3.44	CSID6341
3.	Acetoin	10.97	$C_4H_8O_2$	88.0522	89.06	3.38	CSID21105851
4.	9-Oxononanoic acid	11.44	$C_9H_{16}O_3$	172.10	173.1171	4.53	CSID68222
5.	Piperidine	11.83	$C_5H_{11}N$	85.0886	86.0964	-0.71	CSID7791
6.	Umbellic acid	12.40	$C_{15}H_{22}O$	180.0422	181.05	3.36	CSID393925
7.	L-(+)-Mandelic acid	12.40	$C_{15}H_{22}O$	152.0473	153.0551	3.74	CSID388690
8.	2-Hydroxyphenylacetic acid	12.40	$C_{15}H_{22}O$	152.0473	153.0551	3.74	CSID11476
9.	3-Hydroxyphenylacetic acid	12.40	$C_{15}H_{22}O$	152.0473	153.0551	3.74	CSID11624
10.	4-Hydroxyphenylacetic acid	12.40	$C_9H_8O_4$	152.0473	153.0551	3.74	CSID124
11.	3,4-Dihydroxyphenylacetaldehyde	12.40	$C_9H_8O_4$	152.0473	153.0551	3.74	CSID106504
12.	Vanillin	12.40	$C_{23}H_{24}O_8$	152.0473	153.0551	3.74	CSID13860434
13.	Artemisinic aldehyde	12.48	$C_{15}H_{22}O$	218.0372	219.045	0.10	CSID13115339
14.	Zerumbone	12.48	$C_{15}H_{22}O$	218.0372	219.045	0.10	CSID13450367
15.	Solavetivone	12.48	$C_{15}H_{22}O$	218.0372	219.045	0.10	CSID390842
16.	5',5'-Dimethyl-1,1'-bi(cyclohexane)-1',3-diene-4-carbaldehyde	12.48	$C_{15}H_{22}O$	218.0372	219.045	0.10	CSID58829872
17.	4-Hydroxyphenylpyruvate	12.95	$C_9H_8O_4$	180.0424	181.0502	4.38	CSID954
18.	Trans-caffeic acid	12.95	$C_9H_8O_4$	180.0424	181.0502	4.38	CSID600426
19.	Butyl propionate	18.50	$C_7H_{14}O_2$	130.0993	131.1071	4.28	CSID11045
20.	(9E)-9-Octadecenedioate	19.5	$C_{18}H_{30}O_4$	310.2148	311.2226	-0.33	CSID15077503
21.	1-Naphthyl β-D-glucopyranoside	21.36	$C_{16}H_{18}O_6$	307.1189	306.111	4.72	271.0979
Methanol extract							
22.	1,2-Dihydropyrimidine	12.14	$C_4H_6N_2$	82.0526	83.0604	0.49	CSID1027
23.	2-(2-Methylenecyclopropyl)-3-oxosuccinic acid	20.55	$C_8H_8O_5$	184.0373	185.0451	4.15	CSID81407866
24.	Furaneol	27.02	$C_6H_8O_3$	128.0202	129.028	1.61	CSID18218
25.	Ketopantolactone	27.02	$C_6H_8O_3$	128.0202	129.028	1.61	CSID38

MM = Monoisotopic mass

Table 4 — ICP-MS analysis of *L. chinensis* seeds

S.No.	Elements	Sample concentration (ppm)
1	Fe	15.95
2	Zn	10.86
3	As	bdl
4	Co	bdl
5	Cu	3.37
6	Mn	2.97
7	Ni	0.16
8	Pb	0.31
9	Rb	3.38
10	Se	0.09
11	Sr	5.61
12	U	0.06
13	V	bdl
14	Ca	77.09
15	Al	0.37

bdl: below detection limit

Table 5 — Fatty acid composition of *L. chinensis* seeds

S.No.	Type of fatty acid	Area %
1.	Caproic acid (C6:0)	0.14
2.	Caprylic acid (C8:0)	0.25
3.	Capric acid (C10:0)	0.28
4.	Lauric acid (C12:0)	11.66
5.	Myristic acid (C14:0)	5.20
6.	Pentadecanoic acid (C15:0)	0.05
7.	Palmitic acid (C16:0)	9.96
8.	Palmitoleic acid (C16:1)	0.55
9.	Heptadecanoic acid (C17:0)	0.19
10.	Cis-10-pentadecanoic acid (C15:1)	0.69
11.	Stearic acid (C18:0)	2.08
12.	Elaidicacid (C18:1 Δ9t)	10.36
13.	Oleic acid (C18:1Δ9c)	18.44
14.	Linolelaidic acid (C18:2Δ6t)	0.50
15.	Linoleic acid (C18:2)	11.30
16.	Arachidic acid (C20:0)	0.61
17.	Linolenic acid (C18:3)	10.20
18.	Cis-11-eicosenoic acid (C20:1)	0.21
19.	Heniecosanoic acid (C21:0)	0.32
20.	Cis-8,11,14-eicosadienoic acid (C20:3Δ6)	0.53
21.	Behenic acid (C22:0)	0.66
22.	Cis-11,14-eicosadienoic acid (C20:2)	0.36
23.	Erucic acid (C22:1Δ9c)	6.60
24.	Methyl cis-5,8,11,14-eicosatetraenoic acid (C20:4)	0.47
25.	Cis-13,16-docosadienoic acid (C22:2)	4.80
26.	Lignoceric acid (C24:0)	0.46
27.	Nervonic acid (C24:1)	3.15

element with 77.09 ppm concentration in the *L. chinensis* seeds. All the minerals present in *L. chinensis* seeds are extremely beneficial to human health. For example, calcium can be beneficial to build bones, iron metal is required in the production of haemoglobin. Zinc, copper, and manganese are important cofactors in the bone metabolism[26].

FAME analysis

The fatty acid composition of *L. chinensis* seeds is shown in Table 5. The highest quantity of fatty acid was found for oleic acid (18.44%), followed by lauric acid (11.66%), linoleic acid (11.30%), linolenic acid (10.20%), and palmitic acid (9.96%). Oleic acid reduces the risk of cardiovascular disease and improves insulin sensitivity[27]. Banerjee *et al.*[28] has concluded that this fatty acid has a future therapeutic effect associated with human cell apoptosis. It also shows anti-inflammatory and antioxidant properties.

Lauric acid has antimicrobial and anti-viral properties making it a useful natural alternative to conventional antibiotics and antiviral drugs. It contains a number of health benefits, including improved heart health, lower cholesterol level, increased levels of high-density lipoprotein (HDL), and weight loss[29].

Palmitic acid helps in regulating glucose metabolism and the development of insulin resistance in type 2 diabetes[30]. Linoleic acid that cannot be synthesised by the human body is an essential omega-6 PUFA. It has various biological effects over human body cells such as cell membrane flexibility, reduced cardiovascular disease, lower cholesterol levels, immune system function, and inflammation regulation[31]. Another essential omega-3 polyunsaturated fatty acid, linolenic acid is an important component of cell membranes involved in many biological processes in the body. It shows anti-inflammatory effects and prevention as well treatment of chronic inflammatory diseases[32].

Antioxidant activity

DPPH radical scavenging assay

The anti-oxidant activity of methanolic extract of *L. chinensis* seeds was examined by DPPH radical scavenging test. DPPH radical assay is widely used for screening antioxidants for the majority of the plant products and plant extract's activity to quench free radicals. The antioxidants present in the extract, decolourise DPPH radical, this free radical decolourisation follows the mechanism of electrons or hydrogen atoms to neutralise DPPH[14,20]. The result of the free radical scavenging activity is displayed in Figure 1 which shows that the maximum antioxidant activity was shown by the solution with a sample concentration of 200 µL with a colour change from purple to yellow, confirming interaction of the antioxidants in the sample with the free radicals.

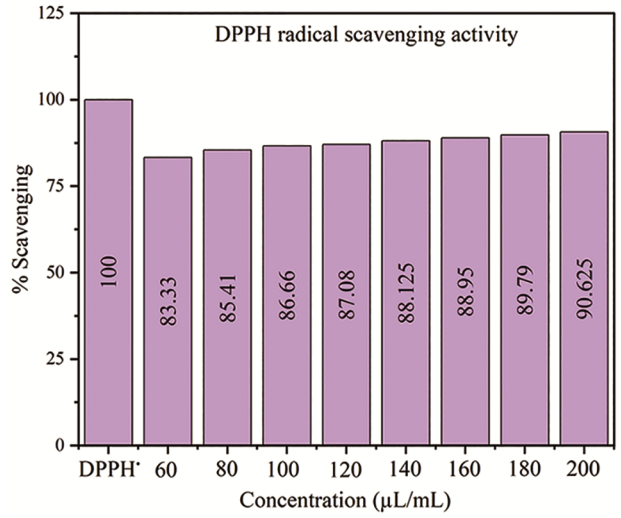


Fig. 1 — Antioxidant activity of methanol extract of *Livistona chinensis* seeds extract

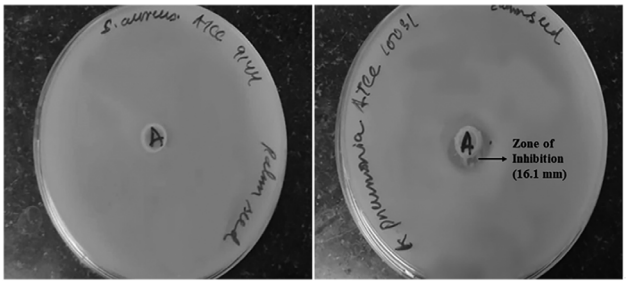


Fig. 2 — Antimicrobial activity of *Livistona chinensis* seeds extract

Antimicrobial activity

The results of the antimicrobial activity revealed that among methanol and aqueous extracts of *L. chinensis,* only the aqueous extract showed the anti-microbial activity against gram-negative bacteria (*Klebsiella pneumoniae*) with inhibition of 16.1 mm while no activity was observed for the same extract for gram-positive bacteria (*Staphylococcus aureus*) as shown in Figure 2. *L. chinensis* inhibits the bacteria growth. From the above results, we can conclude that the aqueous extract of *L. chinensis* could be a source of antimicrobial agent in cosmetic and pharmaceutical formulation.

Conclusion

This study focused on the extraction and exploring the significance of the phytoconstituent of the Indian variety of *L. chinensis* seeds previous studies had shown their effectiveness in various theraputic properties like anti-caner, anti-microbial, and anti-oxidant agent. Our study investigated the presence of significant amount of flavonoids, alkaloids, and phenolic compounds, which possess various biological activities, while proteins, anthocyanins, and phlobatannins were absent. GC-MS and UHPLC-QTOF-MS analysis confirmed the presence of varied classes of phytoconstituents that may have therapeutic application for the treatment of various diseases. The presence of a high percentage of fatty acids in the seeds extract such as oleic acid, linoleic acid, and palmitic acid make *L. chinensis* application more prominent. The methanolic extract showed significant antioxidant activity. The best antimicrobial activity was only shown by aqueous extract of the seeds with inhibition of 16.1 mm. The above results of phytochemicals examination, antioxidant, and antimicrobial profiling of *L. chinensis* seeds concluded that the seeds may be beneficial for new researchers for further research on the application of investigated phytocompounds.

Acknowledgment

The authors would like to thank the Department of Applied Chemistry, Delhi Technological University, for the support and constant encouragement to carry out this work. We are also thankful to Dr. Manoj Pandey (Indian Pharmacopoeia Commission, India), Mr. Hitesh Shirmal (Waters, India) and Mr. Anoop Patyal (Panjab University, India) for helpful discussion.

Conflict of Interest

The authors declare that there is no competing or conflict of interest.

Author Contributions

TB: Methodology, writing- original draft; RS: Conceptualization, review, editing, data analysis; RKG: Conceptualization, supervision, review & editing.

References

1 Khursheed T, Fatima T, Qadri T, Rafiq A, Malik A, *et al.,* Biochemical, nutraceutical and phytochemical characterization of chia and basil seeds: A comparative study, *Int J Food Prop*, 26 (1) (2023) 1-13.

2 Anon, Evaluation of the proximate properties of the seed and physicochemical properties of the oil of Chinese fan palm (*Livistona chinensis*), 1 (4) (2012) 295-303.

3 Zeng X, Wang Y, Qiu Q, Jiang C, Jing Y, *et al.,* Bioactive phenolics from the fruits of *Livistona chinensis*, *Fitoterapia*, 83 (1) (2012) 104-109.

4 Wang Y, Zhai J, Yang D, Han N, Liu Z, *et al.,* Antioxidant, anti-inflammatory, and antidiabetic activities of bioactive compounds from the fruits of *Livistona chinensis* based on network pharmacology prediction, *Oxid Med Cell Longev*, 2021 (10) (2021) 7807046.

5 Wu M, Wang C, Mai C, Chen J, Lai X, *et al.,* Flavonoids from *Livistona chinensis* fruit ameliorates LPS/D-GalN-induced acute liver injury by inhibiting oxidative stress and inflammation, *J Funct Foods*, 61 (3) (2019) 103460.

6 Kaur G & Singh R P, Antibacterial and membrane damaging activity of *Livistona chinensis* fruit extract, *Food Chem Toxicol*, 46 (7) (2008) 2429-2434.

7 Singh R P & Kaur G, Hemolytic activity of aqueous extract of *Livistona chinensis* fruits, *Food Chem Toxicol*, 46 (2) (2008) 553-556.

8 Giri B S, Sonwani R K, Varjani S, Chaurasia D, Varadavenkatesan T, *et al.,* Highly efficient bio-adsorption of malachite green using Chinese fan-palm biochar (*Livistona chinensis*), *Chemosphere*, 287 (2022) 132282.

9 Hasan N, Ariffin S, Azzeme A M, Hasbullah N I, Nawahwi M Z, *et al.,* Preliminary phytochemical screening of medicinal herb, SAMBAU PAYA (*Chloranthus erectus*), *Mater Today Proc*, 12 (2023) 3956.

10 Hailu T, Gupta R K & Rani A, Phytochemicals and antioxidant activity of sisymbrium irio l. Seeds, *Indian J Tradit Know*, 20 (1) (2021) 26-32.

11 Sofowora A, Research on medicinal plants and traditional medicine in Africa, *J Altern Complement Med*, 2 (3) 1996 365-372.

12 Said Nasser Al-Owamri F, Saleh Abdullah Al Sibay L, Hemadri Reddy S, Althaf Hussain S & Subba Reddy Gangireddygari V, Phytochemical, antioxidant, hair growth and wound healing property of *Juniperus excelsa*, *Olea oleaster* and *Olea europaea*, *J King Saud Univ - Sci*, 35 (2) (2023) 102446.

13 Lachman J, Hejtmánková A, Táborský J, Kotíková Z, Pivec V, *et al.,* Evaluation of oil content and fatty acid composition in the seed of grapevine varieties, *LWT - Food Sci Technol*, 63 (1) (2015) 620-625.

14 Gaur P, Bhatia S, Andola H C & Gupta R K, *In vitro* radical scavenging activity and antimicrobial potential of *Berberis asiatica Roxb*. ex DC, fruit extracts in four different processed forms, *Indian J Tradit Know*, 16 (4) (2017) 706-713.

15 Nandagoapalan V, Doss A & Marimuthu C, Phytochemical analysis of some traditional medicinal plants, *Biosci Disc,* 7 (1) (2016) 17-20.

16 Wadood A, Phytochemical analysis of medicinal plants occurring in local area of mardan, *Biochem Anal Biochem*, 2 (4) (2013) 2-5.

17 Yadav M, Chatterji S, Gupta S K & Watal G, Preliminary phytochemical screening of six medicinal plants used in traditional medicine, *Int J Pharm Pharm Sci*, 6 (5) (2014) 539-542.

18 Abdel-Aal EI, Haroon A M & Mofeed J, Successive solvent extraction and GC-MS analysis for the evaluation of the phytochemical constituents of the filamentous green alga Spirogyra longata, *Egypt J Aquat Res*, 41 (3) (2015) 233-246.

19 Shan L, Tyagi A, Chen X, Yan P & Oh D H, Potential anti-obesity effect of fermented adzuki beans and their untargeted metabolomics using UHPLC-QTOF-MS, *Food Biosci*, 52 (2023) 102380.

20 Hailu T, Sharma R, Mann S, Gupta P, Gupta R K, *et al.,* Determination of bioactive phytochemicals, antioxidant and anti-inflammatory activity of *Colchicum autumnale* L. (suranjanshireen), *Indian J Nat Prod Resour*, 12 (1) (2021) 52-60.

21 Rahman T U, Khan H, Liaqat W, & Zeb M A, Phytochemical screening, green synthesis of gold nanoparticles, and antibacterial activity using seeds extract of Ricinus communis L, *Microsc Res Tech*, 85 (1) (2022) 202-208.

22 Hifnawy M S, Issaa M Y, El-seedi H, & Ashour R M S, Phytochemical study, nutritional evaluation and, *Jordan J Biol Sci*, 14 (1) (2021) 163-171.

23 Phillips S M & Van Loon L J C, Dietary protein for athletes: From requirements to optimum adaptation, *J Sports Sci*, 29 (sup1) 2011 S29-S38.

24 Cui Z, Wang Z, Zheng M & Chen T, Advances in biological production of acetoin: A comprehensive overview, *Crit Rev Biotechnol*, 42 (8) (2022) 1135-1156.

25 Abdelshaheed M M, Fawzy I M, El-Subbagh H I & Youssef K M, Piperidine nucleus in the field of drug discovery, *Futur J Pharm Sci*, 7 (1) (2021) 188.

26 Saltman P D & Strause L G, The role of trace minerals in osteoporosis, *J Am Coll Nutr*, 12 (4) (1993) 384-389.

27 Waldman H S, Renteria L I & McAllister M J, Time-restricted feeding for the prevention of cardiometabolic diseases in high-stress occupations: A mechanistic review, *Nutr Rev*, 78 (6) (2020) 459-464.

28 Banerjee A, Dey T, Ghosh A K, Mishra S, Bandyopadhyay D, *et al.,* Insights into the ameliorative effect of oleic acid in rejuvenating phenylhydrazine induced oxidative stress mediated morpho-functionally dismantled erythrocytes, *Toxicol Reports*, 7 (11) (2020) 1551-1563.

29 Offor B C, Dubery I A & Piater L A, Prospects of gene knockouts in the functional study of MAMP-triggered immunity: A review, *Int J Mol Sci*, 21 (7) (2020) 2540.

30 Carta G, Murru E, Banni S & Manca C, Palmitic acid: Physiological role, metabolism and nutritional implications, *Front Physiol*, 8 (11) (2017) 1-14.

31 Brenner H & Schöttker B, Vitamin D insufficiency may account for almost nine of ten COVID-19 deaths: time to act. comment on: "Vitamin D deficiency and outcome of COVID-19 patients", *Nutrients*, 12 (12) (2020) 3642.

32 Simopoulos A P, Omega-3 fatty acids in inflammation and autoimmune diseases, *J Am Coll Nutr*, 21 (6) (2002) 495-505.

frontiers | Frontiers in Immunology

TYPE Review
PUBLISHED 21 February 2024
DOI 10.3389/fimmu.2024.1324018

Check for updates

OPEN ACCESS

EDITED BY
Daniel Erny,
University of Freiburg Medical Center, Germany

REVIEWED BY
Miguel A. Ortega,
University of Alcalá, Spain
Okobi Eko Ekpo,
Khalifa University, United Arab Emirates

*CORRESPONDENCE
Saravana Babu Chidambaram
babupublications@gmail.com;
saravanababu.c@jssuni.edu.in
Meena Kishore Sakharkar
meena.sakharkar@usask.ca

RECEIVED 08 November 2023
ACCEPTED 01 February 2024
PUBLISHED 21 February 2024

CITATION
Hediyal TA, Vichitra C, Anand N, Bhaskaran M, Essa SM, Kumar P, Qoronfleh MW, Akbar M, Kaul-Ghanekar R, Mahalakshmi AM, Yang J, Song B-J, Monaghan TM, Sakharkar MK and Chidambaram SB (2024) Protective effects of fecal microbiota transplantation against ischemic stroke and other neurological disorders: an update.
Front. Immunol. 15:1324018.
doi: 10.3389/fimmu.2024.1324018

COPYRIGHT
© 2024 Hediyal, Vichitra, Anand, Bhaskaran, Essa, Kumar, Qoronfleh, Akbar, Kaul-Ghanekar, Mahalakshmi, Yang, Song, Monaghan, Sakharkar and Chidambaram. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Protective effects of fecal microbiota transplantation against ischemic stroke and other neurological disorders: an update

Tousif Ahmed Hediyal[1,2], C. Vichitra[1,2], Nikhilesh Anand[3], Mahendran Bhaskaran[4], Saeefh M. Essa[5], Pravir Kumar[6], M. Walid Qoronfleh[7], Mohammed Akbar[8], Ruchika Kaul-Ghanekar[9], Arehally M. Mahalakshmi[1,2], Jian Yang[10], Byoung-Joon Song[11], Tanya M. Monaghan[12,13], Meena Kishore Sakharkar[10*] and Saravana Babu Chidambaram[1,2*]

[1]Department of Pharmacology, JSS College of Pharmacy, JSS Academy of Higher Education & Research, Mysuru, KA, India, [2]Centre for Experimental Pharmacology and Toxicology, JSS Academy of Higher Education & Research, Mysuru, KA, India, [3]Department of Pharmacology, American University of Antigua, College of Medicine, Saint John's, Antigua and Barbuda, [4]College of Pharmacy and Pharmaceutical Sciences, Frederic and Mary Wolf Centre University of Toledo, Health Science, Toledo, OH, United States, [5]Department of Computer Science, Northwest High School, Bethesda, MD, United States, [6]Molecular Neuroscience and Functional Genomics Laboratory, Department of Biotechnology, Delhi Technological University (Formerly DCE), Delhi, India, [7]Q3CG Research Institute (QRI), Research and Policy Division, Ypsilanti, MI, United States, [8]Division of Neuroscience and Behavior, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, MD, United States, [9]Symbiosis Centre for Research and Innovation (SCRI), Cancer Research Lab, Symbiosis School of Biological Sciences (SSBS), Symbiosis International University (SIU), Pune, Maharashtra, India, [10]Drug Discovery and Development Research Group, College of Pharmacy and Nutrition, University of Saskatchewan, Saskatoon, SK, Canada, [11]Section of Molecular Pharmacology and Toxicology, Laboratory of Membrane Biochemistry and Bio-physics, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Rockville, MD, United States, [12]National Institute for Health Research Nottingham Biomedical Research Centre, University of Nottingham, Nottingham, United Kingdom, [13]Nottingham Digestive Diseases Centre, School of Medicine, University of Nottingham, Nottingham, United Kingdom

The bidirectional communication between the gut and brain or gut-brain axis is regulated by several gut microbes and microbial derived metabolites, such as short-chain fatty acids, trimethylamine N-oxide, and lipopolysaccharides. The Gut microbiota (GM) produce neuroactives, specifically neurotransmitters that modulates local and central neuronal brain functions. An imbalance between intestinal commensals and pathobionts leads to a disruption in the gut microbiota or dysbiosis, which affects intestinal barrier integrity and gut-immune and neuroimmune systems. Currently, fecal microbiota transplantation (FMT) is recommended for the treatment of recurrent *Clostridioides difficile* infection. FMT elicits its action by ameliorating inflammatory responses through the restoration of microbial composition and functionality. Thus, FMT may be a potential therapeutic option in suppressing neuroinflammation in post-stroke conditions and other neurological disorders involving the neuroimmune axis.

Specifically, FMT protects against ischemic injury by decreasing IL-17, IFN-γ, Bax, and increasing Bcl-2 expression. Interestingly, FMT improves cognitive function by lowering amyloid-β accumulation and upregulating synaptic marker (PSD-95, synapsin-1) expression in Alzheimer's disease. In Parkinson's disease, FMT was shown to inhibit the expression of TLR4 and NF-κB. In this review article, we have summarized the potential sources and methods of administration of FMT and its impact on neuroimmune and cognitive functions. We also provide a comprehensive update on the beneficial effects of FMT in various neurological disorders by undertaking a detailed interrogation of the preclinical and clinical published literature.

KEYWORDS

gut microbiota, gut-brain axis, immune cells, neuroimmune axis, neuroinflammation, fecal microbiota transplantation, ischemic stroke, neurological disorders

1 Introduction

The Gut-brain axis (GBA), also known as the microbiota-gut-brain axis, refers to the complex bidirectional connections between gut and brain occurring through the central, autonomic, and enteric nerves systems, endocrine system, and innate and acquired immune systems (1, 2). The gut microbiota (GM) refers to the collection of microorganisms such as bacteria, viruses, protozoa, archaea, and fungi, which are present in the highest density in the colon. The collective genome of the GM in the gastrointestinal tract (GIT) is called the gut microbiome (3). An imbalance between intestinal commensal and pathobionts leads to a disruption in the gut microbiota or gut dysbiosis (GD), which in turn affects gut-immune and neuroimmune systems. The crosstalk between the gut and brain suggests that the GM plays an important role in regulating the metabolism, immune system, and vascular system of the host (4).

1.1 Gut microbiota in host homeostasis

The human GM comprises 10–100 trillion microbes constituting a diverse population belonging to both commensal and opportunistic microbes (5). There are more than 100 bacterial phyla; the major two phyla are *Firmicutes* and *Bacteroidetes*, and minor phyla are *Actinobacteria, Proteobacteria, and Verrucomicrobia* (6). A compositional study on the gut microbiome by Hollister et al. revealed that sex, race/ethnicity, environmental factors, and age contribute to variations in the human GM composition (7). The GM and its metabolites mediate the wellness of the body by regulating immunological, metabolic, and neuroendocrinological functions. The functions of the GM and metabolites are diverse, ranging from supporting the absorption of nutrients and minerals, the synthesis of enzymes, vitamins, and amino acids, and the production of bacterial-derived metabolites such as short-chain fatty acids (SCFAs) (acetate, propionate, valeric acid, and butyrate), phenylacetylglutamine, and trimethylamine N-oxide (TMAO); all of which maintain host homeostasis (8). The GM therefore acts as a defense barrier by mediating colonization resistance (preventing pathogenic colonization) through competing for attachment sites and nutrients, and by secretion and production of antimicrobials (9). Bacterial-derived metabolites, particularly SCFAs, help to modulate the maturation and functions of microglia in the brain, implicating its regulatory potential in modifying neuro-inflammatory responses and thus play a role in neurodegenerative diseases (NDDs) (10). The GM also plays an important role in immunomodulation by altering B-cell activation to produce IgA and Group 3 innate lymphoid cells (immune cells found in barrier sites), gut-associated lymphoid tissues, macrophages, dendritic cells in the lamina propria, and effector and regulatory T cells (11). For example, GD is associated with numerous diseases, including Inflammatory bowel disease (IBD) (12), diabetes (13), Parkinson's disease (PD) (14), Alzheimer's disease (AD) (15), autism (16), multiple sclerosis (17), and stroke (18). These reports indicate a strong involvement of GD in various pathological conditions, particularly altering the inflammatory milieu. GD impacts neurological disorders mainly through altering immune responses. Similarly, use of FMT is shown to improve immune dysregulation in *Clostridioides difficile* infection (CDI) and other inflammatory diseases (19, 20). Herein, we comprehensively demonstrate the mechanistic impact of FMT in various neurological disorder with a central focus on inflammation.

Abbreviations: αSyn- α-synuclein; AIS- Acute ischemic stroke; AD- Alzheimer's Disease; ASD-Autism spectrum disorder; Aβ- Amyloid Beta; BBB- Blood brain barrier; CNS- Central nervous system; CDI- *Clostridioides difficile* infection; DAMP- Damage-associated molecular patterns; ENS- Enteric nervous system; DSS- Dextrean sodium sufate; FMT- Fecal microbiota transplantation; GI-Gastrointestinal; GD- Gut dysbiosis; GM- Gut Microbiota; GBA- Gut-brain axis; IBD- Inflammatory bowel disease; IS- Ischemic stroke; MCAO- Middle cerebral artery occlusion; PD- Parkinson's disease; TMAO-trimethylamine N-oxide; Tg-Transgenic.

1.2 Gut microbiota in major central nervous diseases

Dysregulation of the gut-brain axis has been implicated in various neurological disorders including Alzheimer's disease (AD), autism spectrum disorder (ASD), Parkinson's disease (PD), and stroke. These neurological conditions have been shown to cause changes in the bidirectional link, resulting in the emergence of brain-gut disorders including irritable bowel syndrome (IBS) (21). AD etiology has also been strongly linked to GD and microbial-derived metabolites that are directly linked with phosphorylated tau/Aβ42 accumulation in AD (22). Interestingly, *Helicobacter pylori* infection has been linked to AD pathology. It causes abnormal hyperphosphorylation of tau protein as well as the release of inflammatory mediators and β-amyloids (23). GD has increased recognition in PD wherein the initiation of α-synucleinopathy is produced in the enteric nervous system (ENS) during the early stages of the disease (24). PD patients experience a wide range of clinical symptoms including difficulties in swallowing, drooling, delayed gastric emptying, small intestinal bacterial overgrowth, and constipation (25). Furthermore, gut microbes *belonging to Prevotellaceae, Christensenella, Akkermansia, and Lactobacillus* were previously demonstrated to be as associated with PD (26, 27). On the other hand, several studies have reported the GM composition differences between ASD and healthy children, where studies have reported a reduced abundance of beneficial bacterium *Bifidobacterium*, and an increased abundance of pathobionts such as taxa belonging to Clostridium cluster XVIII and *Escherichia/Shigella* (28, 29). These reports indicate that the GM regulates GBA functions and the pathophysiology of neurological disorders.

1.3 Gut microbiota and the neuroimmune axis

The GM exerts diverse beneficial effects on brain health, the neuroimmune system, and defense against pathogenic infections (30). A well-coordinated inflammatory response within the brain is known as a neuroinflammatory reaction, and it is mediated by a variety of cell populations, including astrocytes and microglia. Astrocytes, microglia, and peripherally derived immune cells are the primary producers of inflammatory mediators such as reactive oxygen species, cytokines, and chemokines (31). Strong experimental evidence suggests that the GM and immune-mediated inflammatory responses interact in a complex, dynamic, and bidirectional manner. Germ-free animals serve as an excellent tool for demonstrating the interaction between the GM on the innate and adaptive immune systems in health and disease states (32).

The innate immune response of the gut mucosa is directed against one or more pathogens and antigens (33). The GM in association with antigen-presenting cells (APCs) protects the body against infection. Also, to maintain immune tolerance to the normal gut flora, the dendritic cells (DCs) of Peyer's patches (lymphoid nodules embedded in the gut wall) produce anti-inflammatory cytokines e.g., interleukin (IL)-10 and TGF-β (34). In the adaptive immune system, CD4+ T cells play an important role. The lamina propria of the intestine contains the majority of the intestinal CD4+ T lymphocytes. Naive CD4+ T cells can develop into one of four main subtypes after being stimulated: T helper 1 (Th1), Th2, Th17, or regulatory T cell (Treg) (35). However, in GD, the adaptive immune system activates CD4+ T cells which further interact with intestinal immune cells such as Th1, Th2, Th17, and Tregs, which lead to neuroinflammation and neuronal death (36).

Pathogenic microbes/microbial components trigger inflammatory reactions via pattern recognition receptor (PRP) families that include pathogen-associated molecular patterns/microbe-associated molecular patterns (PAMPs/MAMPs), or by specific molecular structures released by host cells called danger-associated molecular patterns (DAMPs) (dead or dying cells, cancer cells, toxic, or allergenic structures (37). These stereotyped molecular patterns are recognized by tissue-resident immune cells (mast cells and macrophages) and as a result, they elicit an innate immune response resulting in the increased production of cytokines and chemokines and also induce complement activation (38). Localized dendritic cells and macrophages serve as antigen-presenting cells (APCs). To initiate an adaptive immune response, APCs translocate to tissue-draining lymph nodes where they expose local immune cells to foreign antigens via molecules of the major histocompatibility complex (MHC). A persistent inflammatory response results in the recruitment of circulating leukocytes, particularly T-lymphocytes, which enter the tissue and act as effectors of cellular adaptive immunity (Figure 1).

2 Evidence of gut dysbiosis in neurological disorders

Neurological disorders are closely associated with GD. The factors affecting GD are reduced abundance of beneficial bacteria, excessive proliferation of pathobionts, and a reduction in the overall microbial diversity in the gut (39). Numerous observational and animal studies suggest that the GM plays a vital role in the neuropathogenesis of CNS diseases via alteration in the GBA (40). GD leads to the production of metabolites and cytokines, initiating inflammation, influencing the BBB and brain volume, and potentially acting as pseudo neurotransmitters. These effects contribute to the disruption of brain physiology and neuronal function in various neurological disorders, including those characterized by demyelination or neuropsychiatric manifestations (41–43).

2.1 Cerebral stroke

The American Heart Association/American Stroke Association defines stroke as a "disease that affects the arteries leading to and within the brain". Stroke remains the second leading cause of death globally, with high morbidity, disability, and recurrence rates (44). The World Stroke Organization reported over 12.2 million new

10.3389/fimmu.2024.1324018

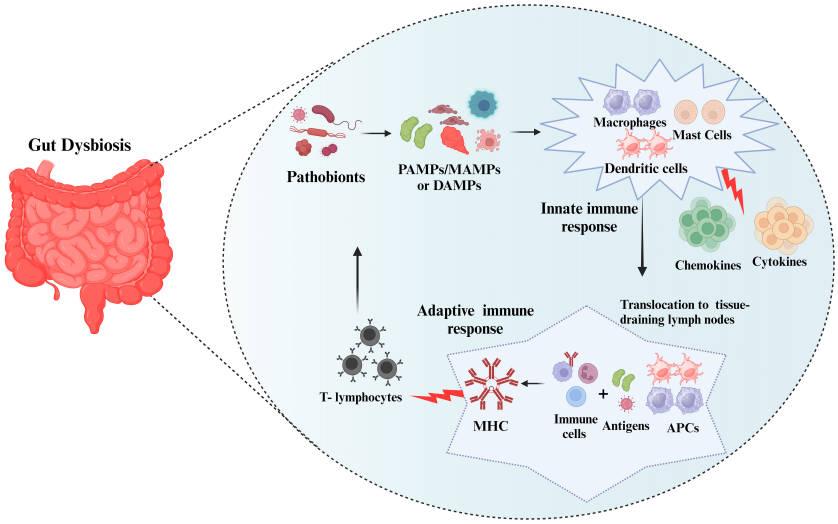


FIGURE 1
Immune response to the pathobionts. The pathobionts trigger inflammatory reactions via pathogen-associated molecular patterns/microbe-associated molecular patterns (PAMPs/MAMPs), or danger-associated molecular patterns (DAMPs). These stereotyped molecular patterns activate the innate immune response (mast cells and macrophages) and adaptive immune response resulting in the increased production inflammatory markers. Subsequently, antigen-presenting cells (APCs) translocate to tissue-draining lymph nodes and form major histocompatibility complex (MHC) and T-lymphocytes will fight against the pathobionts.

stroke cases each year and more than 101 million people that are currently alive have experienced a stroke (45). Globally, 9.6 million cases of ischemic stroke and 4.1 million cases of hemorrhagic stroke have been documented, with 90% of these cases being triggered by modifiable risk factors. In India, stroke is the fourth and fifth leading causes of death and disability, respectively. The occurrence of stroke ranges from 105 to 152/100,000 persons annually in India (46).

In mice following stroke, T lymphocytes migrate from the small intestinal lamina propria or intestinal Peyer's patches to the brain and/or the leptomeninges. Immune cells such as Th1 cells, Th17 cells, monocytes, and toxic chemicals from the gut, and bacteria migrate to the infarct region, which inhibits the migration of anti-inflammatory Treg cells (47). Pathobionts stimulate dendritic cell migration to mesenteric lymph nodes, which promotes T cells to differentiate into Treg cells (48). Impaired Treg cell migration to the lamina propria increases γδT-cell differentiation. The migration of these pro-inflammatory T cells from the lamina propria to the meninges increases the inflammatory milieu and infarct size (49). Additionally, the cytokines released by activated microglia and the DAMPs produced as a result of brain damage stimulate the vagus nerve, which ultimately results in GD (50). The increased translocation of bacterial toxins, pathogenic bacteria, and toxic metabolites produced by these alterations in the GIT subsequently leak additional gut inflammatory and immune cells into the bloodstream, leading to stroke-damaged areas (51). As a result, the outcome of post-stroke treatment is negatively impacted by the inflammatory loop which is triggered by GD. Cheng et al. demonstrated raised plasma levels of LPS and pro-inflammatory cytokines in cynomolgus monkeys post-stroke state, which corresponded with the relative abundance of the phylum *Bacteroidetes* (52). LPS is an immunogenic endotoxin produced by the Gram-negative bacteria (*Yersinia pestis, Klebsiella pneumonia, Pseudomonas aeruginosa, Porphyromonas gingivalis, bifidobacterial species, Chlamydia trachomatis and Francisella tularensis*) that can either directly cause neuroinflammation or peripheral immune cells to migrate into the brain (50).

Ischemic stroke is caused by atherosclerotic plaque or emboli occluding blood vessels, which reduces blood flow and oxygen availability, causing apoptosis and neuronal death (53, 54). Hemorrhagic stroke occurs as a result of deep perforating vasculopathy associated with high blood pressure, amyloid angiopathy of cerebral arteries with microbleeds, and clinical hemorrhages typically affecting the basal ganglia, cerebellum, pons, or thalamus (55). Clinical risk factors for stroke are hypertension, physical inactivity, unhealthy diet, tobacco use, use of alcohol, atrial fibrillation, raised blood lipid levels, genetic disposition, obesity, stress, diabetes, and depression. Post-stroke GI complications such as dysphagia, constipation, fecal incontinence, and GI bleeding are more frequently reported and are characterized by severe inflammation, bowel obstruction, and gut dysbiosis. These complications fuel disease progression and slow down the neuronal recovery of the degenerating brain (56).

Some studies have reported that GM composition influences stroke prognosis (57, 58). The GBA regulates immunological functions in the post-stroke condition while altered GM composition and abundance affect post-stroke outcomes mainly via gut leakiness, endotoxemia, and the neuroimmune axis. These mechanisms allow the dissemination and translocation of resident microflora, cellular, and humoral factors, including metabolites, immune cells, and cytokines/chemokines from the gut to the brain (59).

In ischemic stroke (IS), local and systemic inflammatory responses are augmented (60). Within 24 hours, monocytes (innate response) penetrate the brain, reaching peak numbers 3-5 days after acute ischemic stroke (AIS) (61). Neutrophils play a controversial role in AIS. Neutrophils have been linked to several mechanisms, including the generation of reactive oxygen species and the release of metalloproteinases, but it has also been shown that N2 neutrophils are neuroprotective (62). Additionally, neurotoxic mechanisms (oxidative stress, neuroinflammation, apoptosis, genotoxic activity, and impaired signaling pathways) activate the release of pro-inflammatory cytokines including interleukin-21 from a cluster of differentiating CD4+ T cells within 24 h post-AIS. Th1 and Th17 subpopulations promote neuroinflammation, whereas Tregs act as neuroprotectors by reducing inflammation (63). Moreover, T cells release cytokines and chemokines into the bloodstream which impact the microbiota composition and lead to GD (64). Stroke is frequently reported to be associated with gut dysmotility, increased intestinal barrier permeability, and translocation of microorganisms and a spike in the concentration of metabolites, such as TMAO and lipopolysaccharide (LPS), to enter the bloodstream (65, 66). These alterations further promote the systemic inflammatory response and disease progression and result in a poor prognosis.

In a Japanese cohort study in ischemic stroke patients, the relative abundance of *Atopobium cluster and Lactobacillus ruminis* were elevated and the relative abundance of *Lactobacillus sakeiakei* subgroup decreased and these changes triggered IL-6 detection in the circulation. Furthermore, the decreased concentration of acetic acid was negatively correlated with the levels of glycated hemoglobin and low-density lipoprotein cholesterol. In contrast, elevated fecal valeric acid concentration was positively correlated with the level of high-sensitive C-reactive protein and white blood cell counts. The changes in metabolites associated with GD in patients with IS are correlated with host metabolism and inflammation (67). These observations suggest a strong pathological involvement of GD in the IS condition.

2.2 Alzheimer's disease

As in AD, GD causes altered release of neurotransmitters, which affects the GBA, resulting in behavioral changes, increased anxiety, mood swings, sleep deprivation, and depression. GD is primarily characterized by an elevation in the *Firmicutes/Bacteroidetes* ratio, which has been linked to the accumulation of gut APP from the early stages of AD (68). Alterations in the composition of the GM in APP/PS1 mice were linked to elevated Aβ levels in the CNS as well as impaired memory performances (69). Similarly, in Tg2576 animals, GD, intestinal epithelial barrier dysfunction, and vascular Aβ accumulation in the intestine occur before the development of cerebral Aβ depositions. The occurrence of Aβ deposition is also noted in intestinal autopsies of AD patients (22). Nevertheless, the relationship between GD, intestinal Aβ deposition, and AD onset still needs to be elucidated.

2.3 Parkinson's disease

GD has been postulated to act as a potential cause of PD. Clinical studies have revealed the incidence of GD in PD patients compared with healthy individuals, and both fecal and mucosal GM compositions have been described to change in PD patients (27, 70, 71). Likewise, in the mouse model of PD (Thy1-αSyn transgenic animals), GM exacerbates motor impairments, microglial activation, and α-Syn accumulation. However, in the absence of GM, the severity of motor impairments decreased but did not eliminate (72). In the rat PD model (wild-type Groningen rats), the treatment with dopamine agonists showed improvement in the GD with decreased abundance of pathobionts (*Lachnospiraceae and Prevotellaceae*) and increased abundance of beneficial gut flora (*Lactobacillus and Bifidobacterium*) (73, 74). Hence, the evidence of both clinical and preclinical studies has confirmed that GD plays a crucial role in PD pathology. However, more research is required to study the GM changes which occur in the post-PD phase.

2.4 Autism spectrum disorder

Researchers have reported many reasons for the association of GD with ASD symptoms development. For example, colonization of germ-free mice (GF) with fecal microbiota from ASD children exhibits ASD-like behavior. Mice colonized with microbiota from ASD subjects exhibited increased autistic behavior compared to normal mice (75). ASD patients were shown to have altered GM composition when compared to normal children (76). However, no specific GM species are causally linked to ASD (28, 77). However, in ASD, the data consistently reports alterations in the phyla *Bacteroidetes/Firmicutes* ratios (28, 78). In addition, a few studies have demonstrated an increased abundance of *Candida albicans*, which produce ammonia and other toxins that are likely to cause autism-related behaviors, in the intestines of ASD children compared to non-ASD children (79, 80). Similarly, *S. cerevisiae* stimulates the production of TNF-α and IL-6 by altering immune functions via activation of TLR ligands, and it is thought these immunological mechanisms may play a role in the onset of ASD (81).

3 Fecal microbiota transplantation

Currently, several potential approaches, such as the use of antibiotics as prophylactics, fecal microbiota transplantation (FMT), prebiotics, probiotics, and dietary interventions, have been shown to reverse post-stroke gut GD and improve quality of life (56). FMT is a microbiota-based therapeutic intervention and involves the transfer of fecal matter from a carefully selected donor into the GIT of a recipient to directly modify the recipient's microbial composition and provide a health benefit (82). FMT is shown to be an effective therapy for GI and non-GI diseases, including neurological disorders such as stroke (83). In the United States, FMT is regulated as a biological agent by the Food

and Drug Administration (FDA) with the source of FMT being largely supported by "stool banks" operated by clinical investigators or by OpenBiome. Since the FDA classifies FMT as an investigational new drug (IND), physicians and scientists are required to submit an IND application (84). However, at the European Union level, FMT still occupies a regulatory grey area where it falls to member states to decide how to regulate FMT. For instance, the UK regulates FMT as a medicinal product, where it is recommended for the treatment of recurrent *Clostridioides difficile* infection (CDI) under national guidelines (85). In Italy and Belgium, FMT is regulated as a tissue or cells under their respective national tissue and cell legislation (86). In Australia, all FMT products are regulated as biologicals with the level of regulation varying with the level of external governance and clinical oversight (87, 88). China has formed an FMT consensus expert group under the Committee of Gut Microbiota in Chinese Society of Gastroenterology since 2016 (89). Elsewhere, the regulations regarding FMT are still taking shape (90).

The safety of FMT is an important factor in its implementation. Generally, FMT has been used to treat both GI and non-GI diseases and typically is associated with few or minimal side effects with strict donor screening (91–93) (Table 1). Most short-term risks are mild and known to be related to the delivery method. However, longer-term side effects have not been established (116). In addition, FMT has demonstrated variable degrees of efficacy in treating different GI problems and is the most effective therapy for the treatment of CDI. In the FMT process, stringent donor screening is necessary to avoid the transmission of infectious diseases. However, there is also a theoretical risk of FMT modulating susceptibility for developing conditions or diseases associated with the intestinal microbiota (117, 118). Donors must not have a family history of metabolic, autoimmune, and malignant conditions. Donors are also extensively screened for potential pathogens (85). The preparation of fecal matter involves the mixing of feces with water or normal saline and is followed by the removal of particulate matter through filtration, as shown in (Figure 2).

3.1 FMT delivery methods

FMT aims to treat a disease by restoring phylogenetic diversity and microbiota composition and function. Donor feces can be administered via the upper or lower GIT through different delivery routes (Figure 3) (119). In the upper GIT, FMT administration via capsule is relatively quick, convenient, inexpensive, and technically simple (120). The FMT procedure via the lower GIT necessitates bowel cleansing, followed by recolonizing the entire colon with favorable bacteria. This delivery route avoids the likelihood of aspiration and vomiting but requires a colonoscopy which is an invasive procedure (121). FMT enemas are easy to administer and are inexpensive. However, some patients have reported an aversion to handling stool, which obviously might interfere with the acceptability of fecal enemas (122). Although in the US, the FDA has reported six incidents of serious infections following FMT therapy (123), FMT is considered a safe, effective, and generally well-tolerated therapy with virtually no interim adverse effects if performed correctly. However, the evidence available on long-term safety is limited. Thus, it is crucial to establish clinical protocols that allow clinicians to operate with the highest degree of quality and safety assurance with the fewest possible risks associated with the process.

3.2 Role of FMT on neuroimmune functions

FMT increases intestinal microbial diversity and restores a healthy immune system (124). The potential therapeutic effects of FMT are correlated with decreased GD-induced neuroinflammation (Figure 4). For example, FMT therapy downregulates NLRP3 and IL-1 expression levels, which inhibit the NLRP3 inflammasome's activation and prevents the appearance of pathogenic Th17 cells in a rat model of chronic cerebral hypoperfusion. In the intestinal mucosa, activation (phosphorylation) of Stat3 serves as a master regulator of Th17 cell development by upregulating the production of IL-6, IL-17A, IL-22, and RAR-related orphan receptor gamma (RORγt) (125). FMT treatment reduces the activation of microglial cells and inflammatory gene expression in a mouse model of traumatic brain injury (TBI) (126). Another study found that FMT reduced the proliferative capacity of colonic mucosal T cells in colitic mice compared with animals who did not receive FMT. The frequencies of CD8+ T and CD4+ T cells, which express the cytotoxicity-related molecule CD107a, were also decreased in FMT-treated animals. Indeed, colonic T cells from FMT animals treated showed a decreased pro-inflammatory phenotype (127).

In a clinical study, Jacob et al. performed a single FMT delivery by colonoscopy for active UC patients using a 2-donor fecal microbiota preparation. Post-FMT, mucosal CD4+ T-cell analysis showed a decrease in both Th1 and Treg cells (128). Crothers et al. reported that in a single-center prospective study, post-oral FMT treatment reduced the frequencies of Treg and mucosal-associated invariant T (MAIT) cell populations in the peripheral circulation of ulcerative colitis (UC) subjects (106). Wang et al. carried out a prospective study in UC patients, in which FMT administration significantly reduced the serum levels of IL-1Ra, IL-6, interferon-inducible protein (IP)-10, and epithelial neutrophil-activating peptide (ENA)-78. In addition, granulocyte-colony stimulating factor (G-CSF), vascular cell adhesion molecule (VCAM)-1, and mucosae-associated epithelial chemokine (MEC) levels were significantly decreased in sera from UC patients (129).

4 Evidence on the benefits of FMT in neurological disorders

4.1 FMT in stroke: preclinical evidence and outcomes

In recent years, many animal experiments have been performed to define the relationship between GD associated with metabolic, neuropsychiatric, and sleep disorders, as well as intestinal and

10.3389/fimmu.2024.1324018

TABLE 1 FMT in GI and Non-GI Diseases: preclinical and clinical evidence.

Pre-Clinical Studies					
Disease types	Diseases	Species/ Animal Model	FMT Route	Out Comes	References
GI Disorders	Ulcerative Colitis (UC)	Dextran sodium sulfate (DSS) induced UC in Balb/C mice	Enema	FMT reduced the severity of UC-related inflammation in mice, in which markedly decreased MPO activity, lowered TNF- and IL-1 levels, and enhanced IL-10 levels in colon tissue.	(94)
		GF BalB/c with DSS solution induced UC	Oral gavage	FMT significantly alters the metabolism of DSS-induced mice, and alleviate the symptoms of colitis	(95)
	Clostridioides difficile infection (CDI)	C57BL/6 mice induced with cefoperazone and 103 spores of *C. difficile* strain 630	Oral gavage	Conversely, mice who were not given FMT persistently colonized with high levels of *C. difficile*, and the GM in these mice persisted at low diversity	(96)
		C57BL/6J mice induced with antibiotics and 1.5 x 105 CFU	Enema	The efficacy of FMT products which have been frozen or lyophilized in treating *C. difficile* infection	(97)
Non-GI Disorders	Parkinson's Disease	C57BL/6 mice induced PD with MPTP (i.p injection)	Oral gavage	FMT improved physical impairment in PD mice, lowered fecal SCFAs, reduced GM dysbiosis, boosted striatal DA and 5-HT content, reduced microglia and astrocyte activation in the substantia nigra, and decreased expression of TLR4/TNF-α signaling pathway components in the gut and brain.	(98)
		C57BL/6 mice induced PD with MPTP	Oral gavage	In substantia nigra, FMT suppressed the microglial and astrocytes activation. In addition, it downregulated the expression of GSK3β, IL-1β, inducible nitric oxide synthase and phosphorylated PTEN	(99)
	Alzheimer's Disease	Male APPswe/ PSEN1dE9 transgenic mice induced AD	Oral gavage	In transgenic mice, FMT therapy alleviated cognitive impairments and lowered amyloid- β (Aβ) brain deposition. It also boosted Synapsin I expression and decreased COX-2 and CD11b levels.	(100)
		5xFAD mice induced AD	Oral gavage	Following a 7-day FMT, significant alterations in amyloid plaque deposition and memory functions were seen in the 5xFAD mouse model of AD.	(101)
	Autism Spectrum Disorder	Sprague Dawley rats induced autism by propionic acid	Oral gavage	The amount of BDNF expression stabilized after FMT therapy restored the balance of fecal *Clostridium spp*.	(102)
		C57BL/6J male mice induced autism by p-cresol	Oral gavage	FMT of p-cresol treated mice to normal mice induces typical ASD behavior and increases p-cresol production. Nevertheless, FMT from healthy mice to p-cresol-treated animals was able to improve the social behavior, excitability of VTA dopamine neurons, and fecal p-cresol levels.	(103)
Clinical Studies					
Gastrointestinal Disorders	Ulcerative Colitis	Randomized clinical trial that included 73 pateints with mild to moderately active UC	Colonoscopy	In this preliminary study, anaerobically produced donor FMT was given for 1 week to patients with mild to moderate UC compared with autologous FMT. This resulted in a higher chance of remission at 8 weeks.	(104)
		A double-blind randomized controlled trial, 75 patients with active UC	Retention enema	FMT induces remission in a significantly increases the proportion of patients with active UC than placebo.	(105)
		A randomized control trial involving subjects with active UC	Colonoscopy for 1 day, then oral administration of frozen capsules	cFMT was related to long-lasting donor-induced changes in the microbial composition in fecal material. A correlation between changes in MAIT cell cytokine production and treatment response was found in cFMT recipients.	(106)
	CDI	Open-label single arm prospective, noncontrolled, pilot-study involving 7	Oral capsules	In patients with active UC, 50 days of daily multidonor FMT capsules reduced F-calprotectin and temporarily improved symptoms and health-related life quality.	(107)

(Continued)

TABLE 1 Continued

Pre-Clinical Studies					
Disease types	Diseases	Species/ Animal Model	FMT Route	Out Comes	References
Clinical Studies					
		patients with active UC			
		Randomized, controlled, double-blind clinical trial, 46 subjects who had 3 or more recurrences of CDI	Colonoscopy	Colonoscopy-administered donor stool appeared to be safe and was more efficacious than autologous FMT averting further CDI episodes.	(108)
		Randomized clinical trial, involving 19 patients with CDI	Nasojejunal tube, superior endoscopy, or colonoscopy	Single FMT infusion was shown to have an efficacy of between 70 and 75 percent, and multiple infusions of the FMT enhanced this efficacy to between 85 and 90 percent.	(109)
Non- GI	PD	Case series involving 6 patients with PD	Colonoscopy	At 6 months, FMT was safe and improved PD's non-motor and motor symptoms, including constipation.	(110)
		Prospective study included 11 PD patients with constipation	Nasoduodenal tube.	FMT significantly reduced both motor and non-motor PD symptomatology.	(111)
		Preliminary study including 15 patients with PD	Colonoscopy and nasal-jejunal tube	In comparison to nasointestinal FMT, the colonic FMT group demonstrated a considerable improvement and prolonged maintenance of efficacy. Also, it can safely decrease both the motor and non-motor symptoms experienced by PD patients.	(112)
	AD	A case report, 82-year-old man	Infusion	MMSE score, memory, cognition, mood, and social skills all improved after FMT.	(113)
	ASD	Open-label clinical trial involving 40 children with ASD	Capsules	FMT significantly altered neurotransmitters in the serum and improved the GI and behavioral indications of ASD patients.	(114)
		Follow up study, 18 patients diagnosed with ASD	Oral and rectal	FMT significantly improved GI and behavioral symptoms and the symptoms of autism improved significantly.	(115)

cardiovascular diseases (130, 131). Emerging evidence from animal studies has shown that manipulating the GM via FMT as a potential therapeutic approach for the treatment of stroke and post-stroke complications. A recent systematic review has summarized the preclinical animal evidence for the beneficial effects of FMT on stroke management (83). Inflammatory processes and altered metabolites have a detrimental impact on the GM in post-stroke conditions (132). In one study, post-stroke recolonization in germ-free mice resulted in increased levels of pro-inflammatory cytokines such as IFN-γ and IL-17, T-cell migration from the intestine to the post-stroke brain, and an aggravation of the lesion volume and functional impairments. In a separate study, FMT was associated with the reversal of GD and stroke outcomes with neuroprotective action (64). FMT in middle cerebral artery occlusion (MCAO) animal models fed a Western-style high-fat diet was found to inhibit apoptosis by lowering the level of Bax (Bcl-2-associated X protein) and cleaved caspase-3 and increasing Bcl-2, thereby attenuating the cerebral ischemic injury possibly via reduction in oxidative stress and brain apoptosis (133). In MCAO aged mice, FMT by gavaging stool from young mice was shown to improve post-stroke recovery and survival, while the converse was observed when FMT was administered from aged mice. In this latter group, FMT worsened post-stroke mortality indicating the role of age in affecting the gut microenvironment (134). Similarly, a gender-derived GM transplantation study revealed the dominance of microbial communities from female mice in increasing the survival rate, protection from brain damage and decreasing the systemic level of inflammatory cytokines in the post-stroke condition, suggesting sex as another critical factor influencing the gut microenvironment (135) (Figure 5, Table 2).

Lactobacillus species have been shown to decrease oxidative stress, neuronal apoptosis, and cerebral infarction volume via inhibition of TLR-4/NF-κB signaling, and prevented barrier dysfunction by repairing epithelial cells in the gut following cerebral ischemia reperfusion-injury (140). In antibiotic-treated C57BL/6 mice, transplantation of gut microbiota from db/db mice treated with sodium butyrate showed a reduction in cerebral infarct volume with better gut barrier and blood-brain barrier

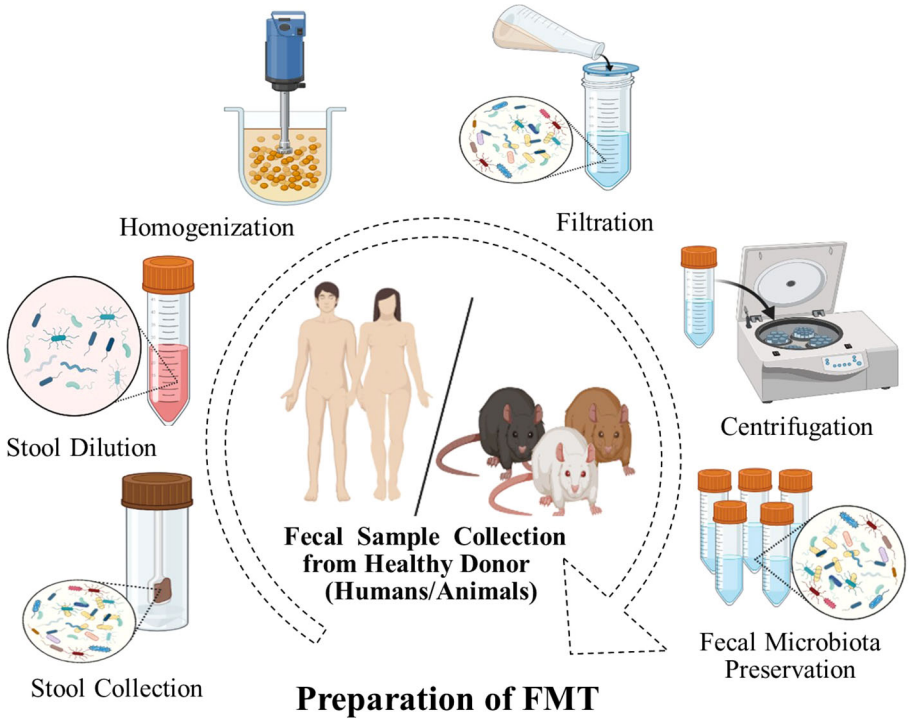


FIGURE 2
FMT preparation. A fecal sample is collected from donors, weighed, and dissolved in a fixed volume of sterile saline and homogenized. The mixture is filtered to remove large particulate matter, which may obstruct the endoscope channel. The filtrate is centrifuged and the supernatant is usually stored at -80°C until required for use.

(BBB) function. These mice also exhibited an increased abundance of butyrate-producing bacteria such as *Ruminococcaceae, Ruminococcu, Lachnospiraceae, and Oscillospira* and showed reduced levels of serum LPS and inflammatory cytokines (IL-6, TNF-α, and IL-1β) in ischemic mice. Furthermore, attenuation of IS injury in the butyrate-treated mice correlated with reduced levels of systemic inflammatory markers such as intercellular adhesion molecule-1 (ICAM-1), vascular cell adhesion molecule-1 (VCAM-1), and matrix metalloproteinase-9 (MMP-9) in the brain. The latter may also be due to the protection of the BBB, marked by the prevention of cerebral capillary endothelial glycocalyx degradation, reflected by the lower levels of syndecan-1 and heparan sulfate in serum (136). FMT therapy improved GM-derived metabolites including SCFAs in the recipient group as compared to the saline group (137). SCFAs support the growth of beneficial bacterial populations such as *Lactobacillus, Butyricicoccus, and Meganonas*. Additionally, FMT also alleviated the level of serum total cholesterol levels, cerebral infarct volume, and intestinal permeability in an MCAO mouse model (137).

Based on these preclinical findings, GD exerts chronic inflammatory responses both peripherally and centrally that accelerate stroke prognosis. Certain bacterial phyla such as *Bacteroidetes, Firmicutes, and Actinobacteria* are altered in the post-stroke condition as depicted by increased F/B ratio (141), Therefore, targeting these bacterial phyla via FMT represents a promising adjuvant therapeutic and prophylactic strategy to protect the brain and gut against stroke-related complications.

4.2 FMT in stroke: clinical evidence and outcomes

The pathophysiology of stroke is strongly linked with gut inflammation and immune responses. This could be a key therapeutic target in post-stroke management. Studies reported that patients with stroke showed altered GM composition compared to healthy subjects, particularly in specific bacterial phyla as mentioned in (Table 3) (145). The GM-derived SCFAs alter brain functions directly or indirectly via immune, vagal, endocrine, and humoral pathways (146). Na Li et al. reported that in ischemic stroke patients, the fecal microbiota composition of SCFA-producing taxa *Odoribacter* and *Akkermansia* was decreased. Additionally, the genus *Christensenellaceae_R-7_group*, *norank_f_Ruminococcaceae*, and Enterobacter levels were correlated to stroke severity, whilst *Christensenellaceae_R-7_group* levels were positively correlated with IS outcomes (147).

In acute ischemic stroke patients, there is an abundance of Enterobacteriaceae and decreased levels of *Fecalibacterium, Parabacteroides Clostridiaceae, and Lachnospira*. Animals receiving FMT from subjects with a high stroke dysbiosis index (SDI) showed a higher abundance of *Enterobacteriaceae* and lower levels of *Lachnospiraceae* which are associated with severe brain injury and poor stroke outcomes, in part displayed by a downward trend of T cells which act as neuroprotectants (143).

In stroke, thrombin causes direct vascular disruption, cellular dysfunction, oxidative stress, and inflammation. In addition, there

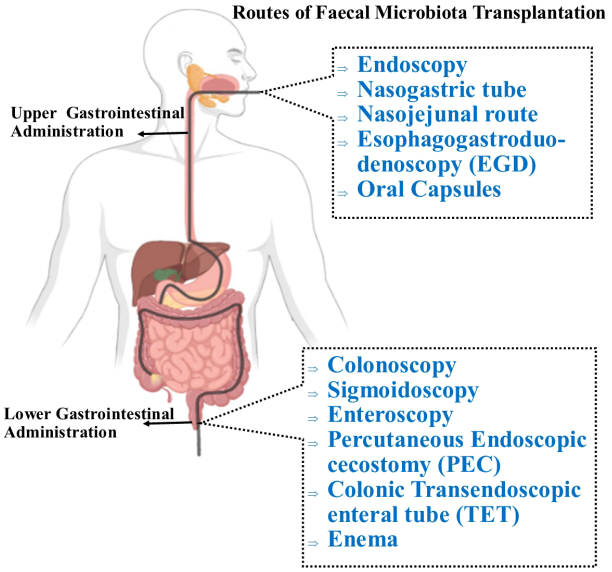


FIGURE 3
Routes of administration for FMT. Through the upper GI route, FMT can be administered via nasogastric tube which is placed through the nose and reaches the stomach or the jejunum (nasojunal tube) to deliver the fecal transplant. The most currently preferred delivery technique involves oral administration of encapsulated FMT, which is both safe and efficient, and is well-tolerated by patients. To deliver the processed fecal matter, the supply tube is inserted into the mouth and passed through the pharynx, oesophagus, stomach, and duodenum (EGD). In the lower GIT, FMT is administered with a colonoscopy, in which the physician inserts a colonoscope into the colon to administer the liquefied donor's stool. Similar to a colonoscopy, a sigmoidoscopy primarily examines and delivers fecal materials to the lower third of the colon. A percutaneous endoscopic cecostomy (PEC) refers to when the C-tube is inserted under endoscopic or imaging guidance to the cecum. FMT can be delivered into the small intestine with esophagogastroduodenoscopy and using a longer conventional endoscope, a double-balloon endoscope (double-balloon enteroscopy) which may be passed through mouth or anus. During endoscopy, the colonic transendoscopic enteral (TET) tube is introduced into the ileocecal junction and connected to the cecum using titanium clips. The first technique to be used was the traditional enema, also referred to as a clyster, which involves injecting the stool preparation into the lower colon via the rectum.

is a direct correlation between thrombin activity in the affected brain hemisphere and the infarction volume (148). Yassene Mohammed et al. reported that the FMT from healthy donors suppressed the onset of thrombin generation in metabolic disorder patients (149). However, the effect of FMT on delaying the onset of thrombin generation in stroke patients is yet to be investigated.

Despite the different aetiologies of neurological disorders, neuronal damage is commonly associated with chronic activation of an innate immune response in the central nervous system (CNS). The GM composition differs in patients compared to healthy controls, indicating a pathophysiological role of the gut microbiome in CNS functions. However, GM restoration by FMT could be an effective treatment approach for several neurological disorders despite the available limited evidence (113).

4.3 Alzheimer's disease

Alzheimer's Disease (AD) is characterized by the accumulation of specific proteins inside or outside cells such as misfolded amyloid-β (Aβ) and tau hyperphosphorylation which forms neurofibrillary tangles. The GM plays a vital role in brain functions including myelinization, neurogenesis, and microglial activation which are closely related to behavioral, mood, and cognitive modulations (130). Mounting evidence indicates that AD is associated with abnormal GM (150). The abundance of *Desulfovibrionaceae* and *Helicobacteraceae* at the family level and *Helicobacter* and *Odoribacter* at the genus level were significantly increased in APPswe/PS1dE9 transgenic (Tg) mice compared with wild-type (WT) mice (69). In an APPswe/PS1dE9 Tg mouse model, FMT improved cognition as evidenced by increased synaptic markers (PSD-95, synapsin-1) and decreased Aβ accumulation and neuroinflammatory markers (COX-2, CD11b). FMT treatment enriched bacterial like *Proteobacteria, Verrucomicrobia*, at phylum levels and *Desulfovibrio and Akkermansia* at genus levels and decreased the abundance of *Bacteroidetes* at phylum levels *and Alloprevotella* at genus levels in the Tg animals (100). Furthermore, FMT restored the levels of *Acidobacteria* and *Bifidobacterium* which significantly delayed AD progression in 3xTg-AD mice (151). Dodiya et al. reported that FMT between age- and sex-matched AD male mice without antibiotics and antibiotic-treated AD male mice partially restored microglial morphology and recolonized intestinal bacteria. Post-FMT in antibiotic-treated male mice, anti-inflammatory cytokines such as IL-10 were expressed more frequently, while pro-inflammatory cytokines like IL-1β, IL-2, IL-3, IL-17A, LIX (CXC5), RANTES (CCL5) and a cluster of differentiation 30 (CD30) and CD40 were reduced (152). Elangovan et al. demonstrated the 5XFAD mouse model showed significantly improved novel object recognition and spatial memory and reduction in amyloid pathology after 7 days of treatment of FMT from wild-type donor mice (101). These data demonstrate that FMT is an effective treatment for AD-like disorders, but further studies are required.

4.4 Parkinson's disease

PD is a chronic, progressive, neurodegenerative condition characterized by accumulation of the presynaptic neuronal protein α-synuclein (αSyn) and dopaminergic neuron loss (110). It is not only associated with motor and non-motor deficits but also affects GI function causing altered bowel movement patterns and abdominal bloating (153). Several studies have reported a relationship between Toll-like receptors (TRLs), particularly TRL4, GM, and αSyn pathology. TLR4 depletion inhibited upward regulation of AP-1 in dopaminergic neurons in the substantia nigra of PD mice, suggesting that TLR4 may play a potential role in PD (154). TLRs have also been demonstrated to activate several signaling pathways, including PI3K/AKT/GSK3 and NF-κB. Additionally, TLR4 activation stimulates microglial cells

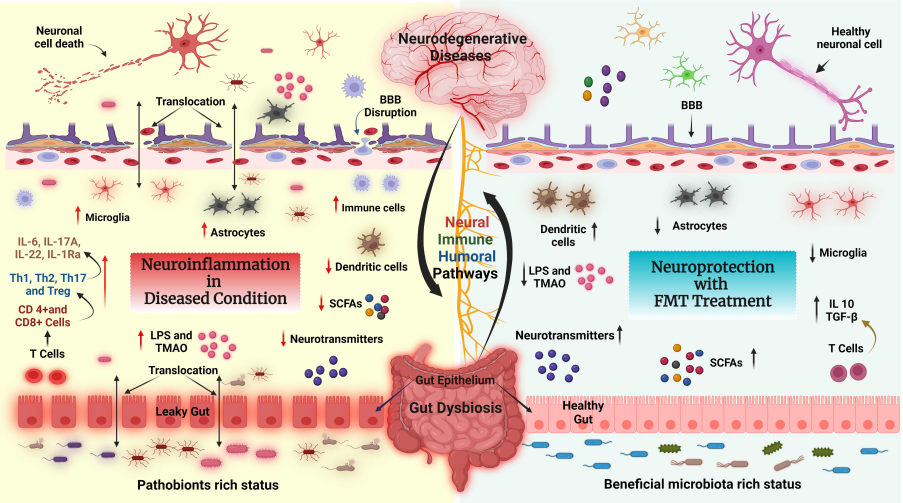


FIGURE 4
FMT modulates the neuroimmune axis. Effect of FMT on the immune system and metabolites in the context of neurological disease and gut dysbiosis (GD) state via the GBA. Left side: Pathobionts induce an inflammatory state during the GD condition. In GD, the adaptive immune system activates CD4+ T cells which further interact with intestinal immune cells [T helper 1 (Th1), Th2, Th17, or regulatory T cell (Treg)] and produce inflammatory cytokines such as IL-6, IL17A, IL22, IL-1Ra, LPS and TMAO, which leads to neuroinflammation and neural cell death. In addition, GD reduces the levels of fecal SCFAs, causing BBB disruption, activating astrocytes and microglia, and inducing neuroinflammation followed by neuronal cell death. Right side: The recolonized GM (beneficial GM) by FMT promotes the production of neurotransmitters, SCFAs, and regulatory T cells through interactions with intestinal immune cells (dendritic cells). Furthermore, FMT restores gut barrier integrity, decreases the production and translocation of TMAO and LPS to the periphery, reduces BBB disruption and activation of brain immune cells, and improves brain functions.

and α-syn, which in turn enhances the synthesis of TNF-α and nuclear translocation of NF-κB (155).

Studies have reported that patients with PD have an increased abundance of genera including *Akkermansia, Lactobacillus, Bifidobacterium, Enterobacteriaceae*, and decreased levels of *Blautia, Faecalibacterium,* and *Prevotella* compared to healthy controls (14, 156). Very recent reports also reveal that FMT treatment improved the motor and non-motor functions in PD patients and reduced the symptoms of constipation (110, 157). FMT treatment inhibited the expression of p-PI3K, p-AKT, TLR4, and NF-κB, as well as TNF-α, confirming a close link between the TLR4/PI3K/AKT/NF-κB signaling and gut microbial dysbiosis in PD (158).

4.5 Autism spectrum disorder

Autism spectrum disorder (ASD) are severe brain or neurodevelopmental disorder that is characterized by deficits in social communication with restricted, repetitive, and stereotyped behaviors that can vary in individuals along a continuum of severity. The pathophysiology of ASD is significantly influenced by the neuroimmune system's involvement and dysregulation, which stimulate microglia, astrocytes, and the release of pro-inflammatory cytokines (159). Inga J´acome et al. reported that in patients (3 to 9-year-old children) with ASD, concentrations of pro-inflammatory cytokines such as IL-1β, IL-6, IL-17, IL-12p40, and TNF-α in plasma were high and correlated with disease severity and progression in comparison with healthy subjects (160). Additionally, children with autism have also shown signs of immune system activation, such as an aberrant CD4:CD8 T cell ratio, a high proportion of DR+ (activated) T cells, elevated neopterin levels in the urine, and increased cytokine production (161). The systemic and neuroinflammatory effects of GM are anticipated to have a direct impact on lymphoid cells and the adaptive immune response. In general, microbial components are known to directly interact with antigen-presenting cells, TLRs, differentiated B cells, T cells, and CD4+ T cells (162). The GM induces activation of Th17 and Th17 lymphocytes which increases systemic inflammation and promotes BBB disruption and CNS inflammation (163).

In recent years, emerging research reported the association between ASD and GM, which was also identified as a contributing component to the prognosis of ASD through the GBA (28, 164). Patients with ASD had higher abundance of *Bacteroides, Parabacteroides, Faecalibacterium, Clostridium,* and *Phascolarctobacterium*, according to a metanalysis report. On the other hand, control subjects have higher levels of *Coprococcus* and *Bifidobacterium* (164). In a clinical cohort study of ASD patients, Ning Li et al. found that the levels of the phyla *Verrucomicrobia*, and genera *Eubacterium Coprostanoligenes, Akkermansia, Ruminococcus, Corprococcus,* and *Christensenellaceae* were significantly altered and that of 5-HT transporter (SERT or 5-HTT) or 5-HT levels were elevated. FMT treatment through oral and rectal delivery led to significant improvement (gastrointestinal symptoms and autism-like behaviors) in ASD patients by reducing *Eubacterium* and *Coprostanoligenes*. In addition, FMT decreased the serum levels of neurotransmitters 5-HT and GABA (114). In a

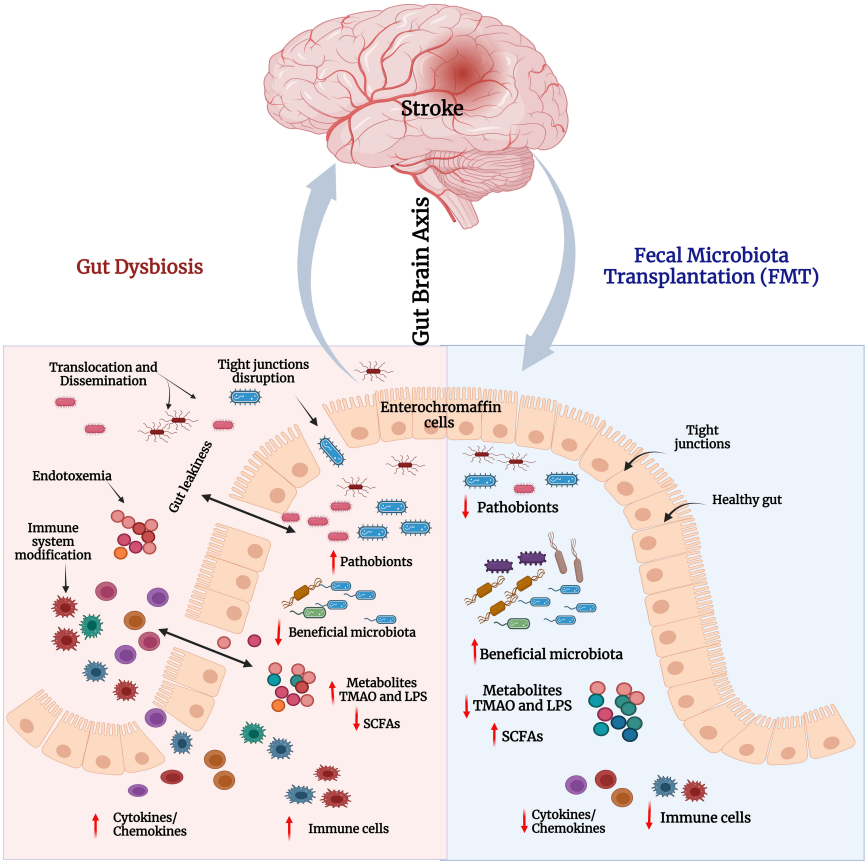


FIGURE 5
Gut dysbiosis and impact of FMT in stroke. Left panel – Stroke-associated gut dysbiosis leads to increased pathobionts and decreased beneficial bacteria in the gut. This leads to decreased levels of short-chain fatty acids (SCFAs), and elevated levels of LPS and TMAO. This triggers immune cells and the secretion of proinflammatory cytokines/chemokines. Increased cytokines and inflammation disrupt the tight junction integrity that induces gut leakiness. Consequently, excessive LPS, TMAO, cytokines, and immune cells pass through loose junctions into the systemic circulation to cause endotoxemia and systemic inflammatory reactions. Right panel - FMT reduces pathobionts and elevates abundance of beneficial bacteria i.e., restores gut microbial homeostasis and gut barrier integrity. Consequently, neuroinflammation and cognitive functions in stroke models would be improved.

single case report, FMT by colposcopy and oral capsule was shown to improve gastrointestinal and ASD-related symptoms and decrease the Childhood Autism Rating Scale (CARS scores) in a patient with ASD (165).

FMT from healthy human GM significantly reduced anxiety-like repetitive behaviors and raised serum levels of chemokines, including GRO-1 (CXCL1), MRIP-1 (CCL3), MCP-3 (CCL7), Eotaxin (CCL11) and RANTES (CCL5) which are crucial for neurogenesis and synaptic transmission in the central nervous system, in ASD mice (166). In another study, FMT from naive wild-type mice significantly reduced levels of TNFα and Iba1 in the mouse brain, normalized *A. muciniphila* abundance to wild-type levels, and restored memory impairments and social disengagement in *Fmr1* knockout mice (167).

The GM affects numerous essential host functions, including the immune response and the nervous system. Both human and animal data do seem to indicate that FMT may improve ASD symptoms. However, as data on the safety profile of FMT and the long-term effects of this treatment in ASD is still limited, further research is needed.

5 Conclusion

In the current review, we have summarized the relationship between GD and the development and prognosis of various neurological diseases including AD, PD, ASD, and stroke. The GM may be considered a potential contributory factor in the pathogenesis of neurological disorders. Restoration of the GM by FMT may attenuate symptoms or progression of neurological disorders mainly via GM-mediated immunological and neural pathways. Neurodegenerative diseases have yet to attract effective disease-altering/modifying therapies, despite decades of intensive research. Current therapies typically try to relieve symptoms but impose severe adverse reactions that restrict usage. Thus, there is an essential need for microbiome-inspired therapeutic alternatives that enhance the quality of life or significantly alter the course of the disease. The information compiled in this review indicates that FMT may be a viable and promising therapeutic option for treating several GI and neurological diseases. Nevertheless, large double-blinded randomized clinical tests (RCTs) are desirable to further elucidate the effect of FMT in numerous disorders of the microbiota GBA.

10.3389/fimmu.2024.1324018

TABLE 2 Preclinical evidence of FMT in stroke outcomes.

Preclinical Evidences			
Study Design	FMT Procedure	Outcomes	References
C57BL/6 male aged and young mice MCAO Stroke model	Oral feeding	FMT from young mice had beneficial effects in MCAO aged mice (improved behavioral functions, cytokine levels, GM levels, infarct volume, SCFAs levels) when compared to aged animal's FMT given to young MCAO mice.	(134)
Mice with MCAO-induced transient focal cerebral ischemia	Oral gavage	FMT improved the BBB, decreased the size of the infarct, improved gut barrier function, and decreased serum levels of LPS, LPS binding protein (LBP), and proinflammatory cytokines.	(136)
Sprague Dawley rats with MCAO induced ischemic stroke	Intragastric	FMT decreased intestinal permeability, T-CHO in serum, neurological impairment, and cerebral infarct volume.	(137)
C57 mice with MCAO induced ischemic stroke	Intragastric	FMT reduced the infarct area, improved behavioral test performance, decreased the level of inflammation, and enhanced the release of beneficial metabolites in female mice compared to male mice in a sex-dependent manner.	(135)
WT C56BL/6J and Rag1-/- mice and GF C56BL/6J and Rag1 -/- mice with MCAO induced cerebral ischemia	Oral gavage	FMT showed neuroprotective effect, reduced lesion size, and regulated T cells in peripheral immune system.	(64)
Mice with MCAO induced ischemic stroke	Oral gavage	Mice who received young donor FMT showed greater post-stroke behavioral improvement and less inflammation in the brain and intestine. Researchers also discovered that the young donor microbiota contained significantly more short-chain fatty acids (SCFAs).	(138)
Sprague Dawley rats with controlled cortical impact model induced TBI	Colonoscopy	After TBI, FMT reduced the TMA levels in the feces and TMAO in serum and ipsilateral and improved neurological deficits possibly via the TMA-TMAO-MsrA signaling pathway.	(139)

There is a lack of molecular understanding, despite clinical and preclinical data supporting the use of FMT in neurological diseases. Most studies did not investigate the possible detrimental effects of FMT, and long-term adverse effects were seldom identified. The efficacy of FMT is influenced by several factors, including donors, antibiotic types, treatment approaches, and microbial composition. To further validate the safety and rule out any potential adverse effects, long-term follow-up and appropriate controls are required. In addition, there has been no uniformity in dose, FMT delivery route, stool filtering technique, frequency of administration, and diet in the studies. Also, there is currently no approved standard for choosing donors for specific disorders, thus more research in this line is necessary. Several studies have shown that FMT improves neurological disorders by improving GI symptoms, and host immune function. However, in future, robust taxonomic resolution methods may be employed for identifying the biomarkers for specific neurological disorders. Finally, further mechanistic studies are warranted to understand the direct neuroprotective effects of specific species or bacterial-derived metabolite(s) on the gut-brain axis using molecular techniques such as human organotypic cultures, synthetic communities and microbiome depleted germ free mouse models to determine causality.

TABLE 3 Preclinical and clinical studies showing alterations in GM composition after stroke.

Mode of the Study	Gut microbiota (Increased abondance)	References
Human cohort study (n= 140)	*Porphyromonadaceae* and *Enterobacteriaceae* and *Lactobacillaceae* and *Akkermansia*	(142)
Human cohort study (n=104) and MCAO study (mice)	*Butyricimonas, Parabacteroides, Rikenellaceae, Ruminococcaceae, Oscillospira, Bilophila, Enterobacteriaceae*	(143)
Mouse MCAO model	*Akkermansia muciniphila* and *clostridial species*	(144)
Human case-control study (n=435)	*Enterobacter, Oscillibacter, Megasphaera,* and *Desulfovibrio*	(145)

Author contributions

TH: Formal analysis, Methodology, Software, Writing – original draft, Investigation, Project administration. CV: Writing – original draft. NA: Writing – original draft. MB: Writing – original draft. SE: Investigation, Writing – original draft. PK: Writing – original draft. MW: Writing – original draft. MA: Writing – original draft. RK-G: Writing – original draft. AM: Writing – original draft. JY: Writing – original draft. BS: Data curation, Investigation, Supervision, Writing – review & editing. TM: Conceptualization, Data curation, Supervision, Writing – review & editing. MS: Conceptualization, Data curation, Supervision, Writing – review & editing. SC: Conceptualization, Data curation, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The work was supported by the "Public Health and Nutrition Division", Department of Biotechnology, Ministry of Science and Technology, Govt of India, (BT/PR38038/PFN/20/1528/2020), and JSS AHER grant (order no. REG/DIR(R)/URG/54/2011-12/5662). Also partly funded by Agricultural Development Fund (ADF) grant from the Provincial Government of Saskatchewan, Canada (422905).

Acknowledgments

The authors thank their respective institutions for providing the required infrastructure to produce this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Madhogaria B, Bhowmik P, Kundu A. Correlation between human gut microbiome and diseases. *Infect Med*. (2022) 1:180–91. doi: 10.1016/j.imj.2022.08.004

2. Ortega MA, Álvarez-Mon MA, García-Montero C, Fraile-Martínez Ó, Monserrat J, Martinez-Rozas L, et al. Microbiota-gut-brain axis mechanisms in the complex network of bipolar disorders: potential clinical implications and translational opportunities. *Mol Psychiatry*. (2023) 28:2645–73. doi: 10.1038/s41380-023-01964-w

3. Gupta S, Allen-Vercoe E, Petrof EO. Fecal microbiota transplantation: in perspective. *Therap Adv Gastroenterol*. (2016) 9:229–39. doi: 10.1177/1756283X15607414

4. Cryan JF, O'Riordan KJ, Cowan CSM, Sandhu KV, Bastiaanssen TFS, Boehme M, et al. The microbiota-gut-brain axis. *Physiol Rev*. (2019) 99:1877–2013. doi: 10.1152/physrev.00018.2018

5. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. (2010) 464:59–65. doi: 10.1038/nature08821

6. Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*. (2017) 550:61–6. doi: 10.1038/nature23889

7. Hollister EB, Gao C, Versalovic J. Compositional and functional features of the gastrointestinal microbiome and their effects on human health. *Gastroenterology*. (2014) 146:1449–58. doi: 10.1053/j.gastro.2014.01.052

8. Dash S, Syed YA, Khan MR. Understanding the role of the gut microbiome in brain development and its association with neurodevelopmental psychiatric disorders. *Front Cell Dev Biol*. (2022) 10:880544. doi: 10.3389/fcell.2022.880544

9. Chung H, Pamp SJ, Hill JA, Surana NK, Edelman SM, Troy EB, et al. Gut immune maturation depends on colonization with a host-specific microbiota. *Cell*. (2012) 149:1578–93. doi: 10.1016/j.cell.2012.04.037

10. Ho L, Ono K, Tsuji M, Mazzola P, Singh R, Pasinetti GM. Protective roles of intestinal microbiota derived short chain fatty acids in Alzheimer's disease-type beta-amyloid neuropathological mechanisms. *Expert Rev Neurother*. (2018) 18:83–90. doi: 10.1080/14737175.2018.1400909

11. Jandhyala SM, Talukdar R, Subramanyam C, Vuyyuru H, Sasikala M, Reddy DN. Role of the normal gut microbiota. *World J Gastroenterol*. (2015) 21:8787–803. doi: 10.3748/wjg.v21.i29.8787

12. Buttó LF, Haller D. Dysbiosis in intestinal inflammation: Cause or consequence. *Int J Med Microbiol*. (2016) 306:302–9. doi: 10.1016/j.ijmm.2016.02.010

13. Karlsson FH, Tremaroli V, Nookaew I, Bergström G, Behre CJ, Fagerberg B, et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature*. (2013) 498:99–103. doi: 10.1038/nature12198

14. Scheperjans F, Aho V, Pereira PAB, Koskinen K, Paulin L, Pekkonen E, et al. Gut microbiota are related to Parkinson's disease and clinical phenotype. *Mov Disord*. (2015) 30:350–8. doi: 10.1002/mds.26069

15. Liu S, Gao J, Zhu M, Liu K, Zhang H-L. Gut microbiota and dysbiosis in alzheimer's disease: implications for pathogenesis and treatment. *Mol Neurobiol*. (2020) 57:5026–43. doi: 10.1007/s12035-020-02073-3

16. Taniya MA, Chung H-J, Al Mamun A, Alam S, Aziz MA, Emon NU, et al. Role of gut microbiome in autism spectrum disorder and its therapeutic regulation. *Front Cell Infection Microbiol*. (2022) 12:915701. doi: 10.3389/fcimb.2022.915701

17. Schepici G, Silvestro S, Bramanti P, Mazzon E. The gut microbiota in multiple sclerosis: an overview of clinical trials. *Cell Transplant*. (2019) 28:1507–27. doi: 10.1177/0963689719873890

18. Yamashiro K, Kurita N, Urabe T, Hattori N. Role of the gut microbiota in stroke pathogenesis and potential therapeutic implications. *ANM*. (2021) 77:36–44. doi: 10.1159/000516398

19. Yadegar A, Pakpoor S, Ibrahim FF, Nabavi-Rad A, Cook L, Walter J, et al. Beneficial effects of fecal microbiota transplantation in recurrent *Clostridioides difficile* infection. *Cell Host Microbe*. (2023) 31:695–711. doi: 10.1016/j.chom.2023.03.019

20. Quraishi MN, Shaheen W, Oo YH, Iqbal TH. Immunological mechanisms underpinning faecal microbiota transplantation for the treatment of inflammatory bowel disease. *Clin Exp Immunol*. (2020) 199:24–38. doi: 10.1111/cei.13397

21. Mayer EA, Ryu HJ, Bhatt RR. The neurobiology of irritable bowel syndrome. *Mol Psychiatry*. (2023) 28:1451–65. doi: 10.1038/s41380-023-01972-w

22. Honarpisheh P, Reynolds CR, Blasco Conesa MP, Moruno Manchon JF, Putluri N, Bhattacharjee MB, et al. Dysregulated gut homeostasis observed prior to the accumulation of the brain amyloid-β in tg2576 mice. *Int J Mol Sci*. (2020) 21:1711. doi: 10.3390/ijms21051711

23. Wang X-L, Zeng J, Yang Y, Xiong Y, Zhang Z-H, Qiu M, et al. Helicobacter pylori filtrate induces alzheimer-like tau hyperphosphorylation by activating glycogen synthase kinase-3β. *J Alzheimer's Dis*. (2015) 43:153–65. doi: 10.3233/JAD-140198

24. Chen M, Mor DE. Gut-to-brain α-synuclein transmission in parkinson's disease: evidence for prion-like mechanisms. *Int J Mol Sci*. (2023) 24:7205. doi: 10.3390/ijms24087205

25. Fasano A, Visanji NP, Liu LWC, Lang AE, Pfeiffer RF. Gastrointestinal dysfunction in Parkinson's disease. *Lancet Neurol*. (2015) 14:625–39. doi: 10.1016/S1474-4422(15)00007-1

26. Petrov VA, Saltykova IV, Zhukova IA, Alifirova VM, Zhukova NG, Dorofeeva YB, et al. Analysis of gut microbiota in patients with parkinson's disease. *Bull Exp Biol Med*. (2017) 162:734–7. doi: 10.1007/s10517-017-3700-7

27. Bedarf JR, Hildebrand F, Coelho LP, Sunagawa S, Bahram M, Goeser F, et al. Functional implications of microbial and viral gut metagenome changes in early stage L-DOPA-naïve Parkinson's disease patients. *Genome Med*. (2017) 9:39. doi: 10.1186/s13073-017-0428-y

28. Strati F, Cavalieri D, Albanese D, De Felice C, Donati C, Hayek J, et al. New evidences on the altered gut microbiota in autism spectrum disorders. *Microbiome*. (2017) 5:24. doi: 10.1186/s40168-017-0242-1

29. West KA, Yin X, Rutherford EM, Wee B, Choi J, Chrisman BS, et al. Multi-angle meta-analysis of the gut microbiome in Autism Spectrum Disorder: a step toward understanding patient subgroups. *Sci Rep*. (2022) 12:17034. doi: 10.1038/s41598-022-21327-9

30. Heiman ML, Greenway FL. A healthy gastrointestinal microbiome is dependent on dietary diversity. *Mol Metab*. (2016) 5:317–20. doi: 10.1016/j.molmet.2016.02.005

31. DiSabato DJ, Quan N, Godbout JP. Neuroinflammation: the devil is in the details. *J Neurochem*. (2016) 139:136–53. doi: 10.1111/jnc.13607

32. Smith K, McCoy KD, Macpherson AJ. Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. *Semin Immunol*. (2007) 19:59–69. doi: 10.1016/j.smim.2006.10.002

33. Thaiss CA, Elinav E. The remedy within: will the microbiome fulfill its therapeutic promise? *J Mol Med*. (2017) 95:1021–7. doi: 10.1007/s00109-017-1563-z

34. Suárez LJ, Arboleda S, Angelov N, Arce RM. Oral versus gastrointestinal mucosal immune niches in homeostasis and allostasis. *Front Immunol*. (2021) 12:705206. doi: 10.3389/fimmu.2021.705206

35. Wu H-J, Wu E. The role of gut microbiota in immune homeostasis and autoimmunity. *Gut Microbes*. (2012) 3:4–14. doi: 10.4161/gmic.19320

36. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. *Cell Res*. (2020) 30:492–506. doi: 10.1038/s41422-020-0332-7

37. Schaefer L. Complexity of danger: the diverse nature of damage-associated molecular patterns *. *J Biol Chem*. (2014) 289:35237–45. doi: 10.1074/jbc.R114.619304

38. Jin HS, Suh H-W, Kim S-J, Jo E-K. Mitochondrial control of innate immunity and inflammation. *Immune Netw*. (2017) 17:77–88. doi: 10.4110/in.2017.17.2.77

39. DeGruttola AK, Low D, Mizoguchi A, Mizoguchi E. Current understanding of dysbiosis in disease in human and animal models. *Inflammatory Bowel Dis*. (2016) 22:1137–50. doi: 10.1097/MIB.0000000000000750

40. Yang D, Zhao D, Ali Shah SZ, Wu W, Lai M, Zhang X, et al. The role of the gut microbiota in the pathogenesis of parkinson's disease. *Front Neurol*. (2019) 10:1155. doi: 10.3389/fneur.2019.01155

41. Bermúdez-Humarán LG, Salinas E, Ortiz GG, Ramirez-Jirano LJ, Morales JA, Bitzer-Quintero OK. From probiotics to psychobiotics: live beneficial bacteria which act on the brain-gut axis. *Nutrients*. (2019) 11:890. doi: 10.3390/nu11040890

42. Guo R, Chen L-H, Xing C, Liu T. Pain regulation by gut microbiota: molecular mechanisms and therapeutic potential. *Br J Anaesthesia*. (2019) 123:637–54. doi: 10.1016/j.bja.2019.07.026

43. Karakan T, Ozkul C, Küpeli Akkol E, Bilici S, Sobarzo-Sánchez E, Capasso R. Gut-brain-microbiota axis: antibiotics and functional gastrointestinal disorders. *Nutrients*. (2021) 13:389. doi: 10.3390/nu13020389

44. Regenhardt RW, Das AS, Ohtomo R, Lo EH, Ayata C, Gurol ME. Pathophysiology of lacunar stroke: history's mysteries and modern interpretations. *J Stroke Cerebrovascular Dis*. (2019) 28:2079–97. doi: 10.1016/j.jstrokecerebrovasdis.2019.05.006

45. Feigin VL, Brainin M, Norrving B, Martins S, Sacco RL, Hacke W, et al. World stroke organization (WSO): global stroke fact sheet 2022. *Int J Stroke*. (2022) 17:18–29. doi: 10.1177/17474930211065917

46. Jones SP, Baqai K, Clegg A, Georgiou R, Harris C, Holland E-J, et al. Stroke in India: A systematic review of the incidence, prevalence, and case fatality. *Int J Stroke*. (2022) 17:132–40. doi: 10.1177/17474930211027834

47. Gelderblom M, Weymar A, Bernreuther C, Velden J, Arunachalam P, Steinbach K, et al. Neutralization of the IL-17 axis diminishes neutrophil invasion and protects from ischemic stroke. *Blood*. (2012) 120:3793–802. doi: 10.1182/blood-2012-02-412726

48. Benakis C, Brea D, Caballero S, Faraco G, Moore J, Murphy M, et al. Commensal microbiota affects ischemic stroke outcome by regulating intestinal γδ T cells. *Nat Med*. (2016) 22:516–23. doi: 10.1038/nm.4068

49. Crapser J, Ritzel R, Verma R, Venna VR, Liu F, Chauhan A, et al. Ischemic stroke induces gut permeability and enhances bacterial translocation leading to sepsis in aged mice. *Aging (Albany NY)*. (2016) 8:1049–63. doi: 10.18632/aging.100952

50. Kalyan M, Tousif AH, Sonali S, Vichitra C, Sunanda T, Praveenraj SS, et al. Role of endogenous lipopolysaccharides in neurological disorders. *Cells*. (2022) 11:4038. doi: 10.3390/cells11244038

51. Tang W, Zhu H, Feng Y, Guo R, Wan D. The impact of gut microbiota disorders on the blood–brain barrier. *Infect Drug Resist*. (2020) 13:3351–63. doi: 10.2147/IDR.S254403

52. Cheng Y-W, Phelps E, Nemes S, Rogers N, Sagi S, Bohm M, et al. Fecal microbiota transplant decreases mortality in patients with refractory severe or fulminant *clostridioides difficile* infection. *Clin Gastroenterol Hepatol*. (2020) 18:2234–2243.e1. doi: 10.1016/j.cgh.2019.12.029

53. Woodruff TM, Thundyil J, Tang S-C, Sobey CG, Taylor SM, Arumugam TV. Pathophysiology, treatment, and animal and cellular models of human ischemic stroke. *Mol Neurodegeneration*. (2011) 6:11. doi: 10.1186/1750-1326-6-11

54. Ortega MA, Fraile-Martinez O, García-Montero C, Haro S, Álvarez-Mon MÁ, De Leon-Oliva D, et al. A comprehensive look at the psychoneuroimmunoendocrinology of spinal cord injury and its progression: mechanisms and clinical opportunities. *Military Med Res*. (2023) 10:26. doi: 10.1186/s40779-023-00461-z

55. Cordonnier C, Demchuk A, Ziai W, Anderson CS. Intracerebral haemorrhage: current approaches to acute management. *Lancet*. (2018) 392:1257–68. doi: 10.1016/S0140-6736(18)31878-6

56. Tuz AA, Hasenberg A, Hermann DM, Gunzer M, Singh V. Ischemic stroke and concomitant gastrointestinal complications- a fatal combination for patient recovery. *Front Immunol*. (2022) 13:1037330. doi: 10.3389/fimmu.2022.1037330

57. Dinan TG, Cryan JF. The microbiome-gut-brain axis in health and disease. *Gastroenterol Clin North Am*. (2017) 46:77–89. doi: 10.1016/j.gtc.2016.09.007

58. Chen J, Chi B, Ma J, Zhang J, Gu Q, Xie H, et al. Gut microbiota signature as predictors of adverse outcomes after acute ischemic stroke in patients with hyperlipidemia. *Front Cell Infect Microbiol*. (2022) 12:1073113. doi: 10.3389/fcimb.2022.1073113

59. Koszewicz M, Jaroch J, Brzecka A, Ejma M, Budrewicz S, Mikhaleva LM, et al. Dysbiosis is one of the risk factor for stroke and cognitive impairment and potential target for treatment. *Pharmacol Res*. (2021) 164:105277. doi: 10.1016/j.phrs.2020.105277

60. Chu HX, Kim HA, Lee S, Moore JP, Chan CT, Vinh A, et al. Immune cell infiltration in Malignant middle cerebral artery infarction: comparison with transient cerebral ischemia. *J Cereb Blood Flow Metab*. (2014) 34:450–9. doi: 10.1038/jcbfm.2013.217

61. Kronenberg G, Uhlemann R, Richter N, Klempin F, Wegner S, Staerck L, et al. Distinguishing features of microglia- and monocyte-derived macrophages after stroke. *Acta Neuropathol*. (2018) 135:551–68. doi: 10.1007/s00401-017-1795-6

62. Cuartero MI, Ballesteros I, Moraga A, Nombela F, Vivancos J, Hamilton JA, et al. N2 neutrophils, novel players in brain inflammation after stroke: modulation by the PPARγ agonist rosiglitazone. *Stroke*. (2013) 44:3498–508. doi: 10.1161/STROKEAHA.113.002470

63. Clarkson BDS, Ling C, Shi Y, Harris MG, Rayasam A, Sun D, et al. T cell–derived interleukin (IL)-21 promotes brain injury following stroke in mice. *J Exp Med*. (2014) 211:595–604. doi: 10.1084/jem.20131377

64. Singh V, Roth S, Llovera G, Sadler R, Garzetti D, Stecher B, et al. Microbiota dysbiosis controls the neuroinflammatory response after stroke. *J Neurosci*. (2016) 36:7428–40. doi: 10.1523/JNEUROSCI.1114-16.2016

65. Praveenraj SS, Sonali S, Anand N, Tousif HA, Vichitra C, Kalyan M, et al. The role of a gut microbial-derived metabolite, trimethylamine N-oxide (TMAO), in neurological disorders. *Mol Neurobiol*. (2022) 59:6684–700. doi: 10.1007/s12035-022-02990-5

66. Zhu W, Romano KA, Li L, Buffa JA, Sangwan N, Prakash P, et al. Gut microbes impact stroke severity via the trimethylamine N-oxide pathway. *Cell Host Microbe*. (2021) 29:1199–1208.e5. doi: 10.1016/j.chom.2021.05.002

67. Chen Y, Liang J, Ouyang F, Chen X, Lu T, Jiang Z, et al. Persistence of gut microbiota dysbiosis and chronic systemic inflammation after cerebral infarction in cynomolgus monkeys. *Front Neurol*. (2019) 10:661. doi: 10.3389/fneur.2019.00661

68. Brandscheid C, Schuck F, Reinhardt S, Schäfer K-H, Pietrzik CU, Grimm M, et al. Altered gut microbiome composition and tryptic activity of the 5xFAD alzheimer's mouse model. *J Alzheimer's Dis*. (2017) 56:775–88. doi: 10.3233/JAD-160926

69. Shen L, Liu L, Ji H-F. Alzheimer's Disease Histological and Behavioral Manifestations in Transgenic Mice Correlate with Specific Gut Microbiome State. *J Alzheimer's Dis*. (2017) 56:385–90. doi: 10.3233/JAD-160884

70. Unger MM, Spiegel J, Dillmann K-U, Grundmann D, Philippeit H, Bürmann J, et al. Short chain fatty acids and gut microbiota differ between patients with Parkinson's disease and age-matched controls. *Parkinsonism Related Disord*. (2016) 32:66–72. doi: 10.1016/j.parkreldis.2016.08.019

71. Pietrucci D, Cerroni R, Unida V, Farcomeni A, Pierantozzi M, Mercuri NB, et al. Dysbiosis of gut microbiota in a selected population of Parkinson's patients. *Parkinsonism Related Disord*. (2019) 65:124–30. doi: 10.1016/j.parkreldis.2019.06.003

72. Sampson TR, Debelius JW, Thron T, Janssen S, Shastri GG, Ilhan ZE, et al. Gut microbiota regulate motor deficits and neuroinflammation in a model of parkinson's disease. *Cell*. (2016) 167:1469–1480.e12. doi: 10.1016/j.cell.2016.11.018

73. van Kessel SP, Bullock A, van Dijk G, El Aidy S. Parkinson's disease medication alters small intestinal motility and microbiota composition in healthy rats. *mSystems*. (2022) 7:e01191–21. doi: 10.1128/msystems.01191-21

74. García-Montero C, Fraile-Martinez O, Rodriguez-Martín S, Saz JV, Rodriguez RA, Moreno JMP, et al. The use of prebiotics from pregnancy and its complications: health for mother and offspring—A narrative review. *Foods*. (2023) 12:1148. doi: 10.3390/foods12061148

75. Sharon G, Cruz NJ, Kang D-W, Gandal MJ, Wang B, Kim Y-M, et al. Human gut microbiota from autism spectrum disorder promote behavioral symptoms in mice. *Cell*. (2019) 177:1600–1618.e17. doi: 10.1016/j.cell.2019.05.004

76. Sivamaruthi BS, Suganthy N, Kesika P, Chaiyasut C. The role of microbiome, dietary supplements, and probiotics in autism spectrum disorder. *Int J Environ Res Public Health*. (2020) 17:2647. doi: 10.3390/ijerph17082647

77. Pulikkan J, Maji A, Dhakan DB, Saxena R, Mohan B, Anto MM, et al. Gut microbial dysbiosis in Indian children with autism spectrum disorders. *Microb Ecol*. (2018) 76:1102–14. doi: 10.1007/s00248-018-1176-2

78. Settanni CR, Bibbò S, Ianiro G, Rinninella E, Cintoni M, Mele MC, et al. Gastrointestinal involvement of autism spectrum disorder: focus on gut microbiota. *Expert Rev Gastroenterol Hepatol*. (2021) 15:599–622. doi: 10.1080/17474124.2021.1869938

79. Iovene MR, Bombace F, Maresca R, Sapone A, Iardino P, Picardi A, et al. Intestinal dysbiosis and yeast isolation in stool of subjects with autism spectrum disorders. *Mycopathologia*. (2017) 182:349–63. doi: 10.1007/s11046-016-0068-6

80. Kantarcioglu AS, Kiraz N, Aydin A. Microbiota–gut–brain axis: yeast species isolated from stool samples of children with suspected or diagnosed autism spectrum disorders and *in vitro* susceptibility against nystatin and fluconazole. *Mycopathologia*. (2016) 181:1–7. doi: 10.1007/s11046-015-9949-3

81. Zou R, Wang Y, Duan M, Guo M, Zhang Q, Zheng H. Dysbiosis of gut fungal microbiota in children with autism spectrum disorders. *J Autism Dev Disord*. (2021) 51:267–75. doi: 10.1007/s10803-020-04543-y

82. Smits LP, Bouter KEC, de Vos WM, Borody TJ, Nieuwdorp M. Therapeutic potential of fecal microbiota transplantation. *Gastroenterology*. (2013) 145:946–53. doi: 10.1053/j.gastro.2013.08.058

83. Pasokh A, Farzipour M, Mahmoudi J, Sadigh-Eteghad S. The effect of fecal microbiota transplantation on stroke outcomes: A systematic review. *J Stroke Cerebrovascular Dis*. (2022) 31:106727. doi: 10.1016/j.jstrokecerebrovasdis.2022.106727

84. Thanush D, Venkatesh MP. Fecal microbiota transplantation: History, procedure and regulatory considerations. *La Presse Médicale*. (2023) 52:104204. doi: 10.1016/j.lpm.2023.104204

85. Mullish BH, Quraishi MN, Segal JP, McCune VL, Baxter M, Marsden GL, et al. The use of faecal microbiota transplant as treatment for recurrent or refractory *Clostridium difficile* infection and other potential indications: joint British Society of Gastroenterology (BSG) and Healthcare Infection Society (HIS) guidelines. *Gut*. (2018) 67:1920–41. doi: 10.1136/gutjnl-2018-316818

86. *Faecal Microbiota Transplantation - EU-IN Horizon Scanning Report*. (2021) 9 (2):e0004221. Available at: https://www.ema.europa.eu/en/documents/report/faecal-microbiota-transplantation-eu-horizon-scanning-report_en.pdf.

87. Haifer C, Kelly CR, Paramsothy S, Andresen D, Papanicolas LE, McKew GL, et al. Australian consensus statements for the regulation, production and use of faecal microbiota transplantation in clinical practice. *Gut*. (2020) 69:801–10. doi: 10.1136/gutjnl-2019-320260

88. Administration (TGA) TG. Faecal Microbiota Transplant (FMT) product regulation | Therapeutic Goods Administration (TGA). (2023). Available at: https://www.tga.gov.au/resources/resource/guidance/faecal-microbiota-transplant-fmt-product-regulation (Accessed January 29, 2024).

89. Shi Y, Yang Y. Fecal microbiota transplantation: Current status and challenges in China. *JGH Open*. (2018) 2:114–6. doi: 10.1002/jgh3.12071

90. Zhang F, Luo W, Shi Y, Fan Z, Ji G. Should we standardize the 1,700-year-old fecal microbiota transplantation? *Off J Am Coll Gastroenterol ACG*. (2012) 107:1755. doi: 10.1038/ajg.2012.251

91. Kumar V, Fischer M. Expert opinion on fecal microbiota transplantation for the treatment of *Clostridioides difficile* infection and beyond. *Expert Opin Biol Ther*. (2020) 20:73–81. doi: 10.1080/14712598.2020.1689952

92. Mazzawi T, Lied GA, Sangnes DA, El-Salhy M, Hov JR, Gilja OH, et al. The kinetics of gut microbial community composition in patients with irritable bowel syndrome following fecal microbiota transplantation. *PloS One*. (2018) 13:e0194904. doi: 10.1371/journal.pone.0194904

93. Choi HH, Cho Y-S. Fecal microbiota transplantation: current applications, effectiveness, and future perspectives. *Clin Endosc*. (2016) 49:257–65. doi: 10.5946/ce.2015.117

94. Zhou J, Zhou Z, Ji P, Ma M, Guo J, Jiang S. Effect of fecal microbiota transplantation on experimental colitis in mice. *Exp Ther Med*. (2019) 17:2581. doi: 10.3892/etm.2019.7263

95. Yang Y, Zheng X, Wang Y, Tan X, Zou H, Feng S, et al. Human fecal microbiota transplantation reduces the susceptibility to dextran sulfate sodium-induced germ-free mouse colitis. *Front Immunol*. (2022) 13:836542. doi: 10.3389/fimmu.2022.836542

96. Seekatz AM, Theriot CM, Molloy CT, Wozniak KL, Bergin IL, Young VB. Fecal microbiota transplantation eliminates *clostridium difficile* in a murine model of relapsing disease. *Infect Immun*. (2015) 83:3838–46. doi: 10.1128/IAI.00459-15

97. Jiang Z-D, Alexander A, Ke S, Valilis EM, Hu S, Li B, et al. Stability and efficacy of frozen and lyophilized fecal microbiota transplant (FMT) product in a mouse model of *Clostridium difficile* infection (CDI). *Anaerobe*. (2017) 48:110–4. doi: 10.1016/j.anaerobe.2017.08.003

98. Sun M-F, Zhu Y-L, Zhou Z-L, Jia X-B, Xu Y-D, Yang Q, et al. Neuroprotective effects of fecal microbiota transplantation on MPTP-induced Parkinson's disease mice: Gut microbiota, glial reaction and TLR4/TNF-α signaling pathway. *Brain Behavior Immun*. (2018) 70:48–60. doi: 10.1016/j.bbi.2018.02.005

99. Zhang T, Wang T, Chen X, Zhao Z, Chen Z. Gut microbiota relieves inflammation in the substantia nigra of chronic Parkinson's disease by protecting the function of dopamine neurons. *Exp Ther Med*. (2022) 23:52. doi: 10.3892/etm.2021.10974

100. Sun J, Xu J, Ling Y, Wang F, Gong T, Yang C, et al. Fecal microbiota transplantation alleviated Alzheimer's disease-like pathogenesis in APP/PS1 transgenic mice. *Transl Psychiatry*. (2019) 9:189. doi: 10.1038/s41398-019-0525-3

101. Elangovan S, Borody TJ, Holsinger RMD. Fecal microbiota transplantation reduces pathology and improves cognition in a mouse model of alzheimer's disease. *Cells*. (2022) 12:119. doi: 10.3390/cells12010119

102. Abuaish S, Al-Otaibi NM, Abujamel TS, Alzahrani SA, Alotaibi SM, AlShawakir YA, et al. Fecal transplant and bifidobacterium treatments modulate gut clostridium bacteria and rescue social impairment and hippocampal BDNF expression in a rodent model of autism. *Brain Sci*. (2021) 11:1038. doi: 10.3390/brainsci11081038

103. Bermudez-Martin P, Becker JAJ, Caramello N, Fernandez SP, Costa-Campos R, Canaguier J, et al. The microbial metabolite p-Cresol induces autistic-like behaviors in mice by remodeling the gut microbiota. *Microbiome*. (2021) 9:157. doi: 10.1186/s40168-021-01103-z

104. Costello SP, Hughes PA, Waters O, Bryant RV, Vincent AD, Blatchford P, et al. Effect of fecal microbiota transplantation on 8-week remission in patients with ulcerative colitis: A randomized clinical trial. *JAMA*. (2019) 321:156–64. doi: 10.1001/jama.2018.20046

105. Moayyedi P, Surette MG, Kim PT, Libertucci J, Wolfe M, Onischi C, et al. Fecal microbiota transplantation induces remission in patients with active ulcerative colitis in a randomized controlled trial. *Gastroenterology*. (2015) 149:102–109.e6. doi: 10.1053/j.gastro.2015.04.001

106. Crothers JW, Chu ND, Nguyen LTT, Phillips M, Collins C, Fortner K, et al. Daily, oral FMT for long-term maintenance therapy in ulcerative colitis: results of a single-center, prospective, randomized pilot study. *BMC Gastroenterol*. (2021) 21:281. doi: 10.1186/s12876-021-01856-9

107. Cold F, Browne PD, Günther S, Halkjaer SI, Petersen AM, Al-Gibouri Z, et al. Multidonor FMT capsules improve symptoms and decrease fecal calprotectin in ulcerative colitis patients while treated – an open-label pilot study. *Scandinavian J Gastroenterol*. (2019) 54:289–96. doi: 10.1080/00365521.2019.1585939

108. Kelly CR, Khoruts A, Staley C, Sadowsky MJ, Abd M, Alani M, et al. Effect of fecal microbiota transplantation on recurrence in multiply recurrent *clostridium difficile* infection. *Ann Intern Med*. (2016) 165:609–16. doi: 10.7326/M16-0271

109. Camacho-Ortiz A, Gutiérrez-Delgado EM, Garcia-Mazcorro JF, Mendoza-Olazarán S, Martínez-Meléndez A, Palau-Davila L, et al. Randomized clinical trial to evaluate the effect of fecal microbiota transplant for initial *Clostridium difficile* infection in intestinal microbiome. *PloS One*. (2017) 12:e0189768. doi: 10.1371/journal.pone.0189768

110. Segal A, Zlotnik Y, Moyal-Atias K, Abuhasira R, Ifergane G. Fecal microbiota transplant as a potential treatment for Parkinson's disease – A case series. *Clin Neurol Neurosurg*. (2021) 207:106791. doi: 10.1016/j.clineuro.2021.106791

111. Kuai X, Yao X, Xu L, Zhou Y, Zhang L, Liu Y, et al. Evaluation of fecal microbiota transplantation in Parkinson's disease patients with constipation. *Microbial Cell Factories*. (2021) 20:98. doi: 10.1186/s12934-021-01589-0

112. Xue L-J, Yang X-Z, Tong Q, Shen P, Ma S-J, Wu S-N, et al. Fecal microbiota transplantation therapy for Parkinson's disease: A preliminary study. *Medicine*. (2020) 99:e22035. doi: 10.1097/MD.0000000000022035

113. Hazan S. Rapid improvement in Alzheimer's disease symptoms following fecal microbiota transplantation: a case report. *J Int Med Res*. (2020) 48:300060520925930. doi: 10.1177/0300060520925930

114. Li N, Chen H, Cheng Y, Xu F, Ruan G, Ying S, et al. Fecal microbiota transplantation relieves gastrointestinal and autism symptoms by improving the gut microbiota in an open-label study. *Front Cell Infection Microbiol*. (2021) 11:759435. doi: 10.3389/fcimb.2021.759435

115. Kang D-W, Adams JB, Coleman DM, Pollard EL, Maldonado J, McDonough-Means S, et al. Long-term benefit of Microbiota Transfer Therapy on autism symptoms and gut microbiota. *Sci Rep*. (2019) 9:5821. doi: 10.1038/s41598-019-42183-0

116. Park S-Y, Seo GS. Fecal microbiota transplantation: is it safe? *Clin Endosc*. (2021) 54:157–60. doi: 10.5946/ce.2021.072

117. Aràjol C, Aira Gómez A, González-Suárez B, Casals-Pascual C, Martí Martí S, Domínguez Luzón MÁ, et al. Donor selection for faecal microbiota transplantation. Consensus document of the Catalan Society of Gastroenterology and the Catalan Society of Infectious Diseases and Clinical Microbiology. *Gastroenterología Y Hepatología (English Edition)*. (2021) 44:175–80. doi: 10.1016/j.gastre.2020.07.005

118. Duvallet C, Zellmer C, Panchal P, Budree S, Osman M, Alm EJ. Framework for rational donor selection in fecal microbiota transplant clinical trials. *PloS One*. (2019) 14:e0222881. doi: 10.1371/journal.pone.0222881

119. Hirsch BE, Saraiya N, Poeth K, Schwartz RM, Epstein ME, Honig G. Effectiveness of fecal-derived microbiota transfer using orally administered capsules for recurrent *Clostridium difficile* infection. *BMC Infect Dis*. (2015) 15:191. doi: 10.1186/s12879-015-0930-z

120. Youngster I, Russell GH, Pindar C, Ziv-Baran T, Sauk J, Hohmann EL. Oral, capsulized, frozen fecal microbiota transplantation for relapsing *Clostridium difficile* infection. *JAMA*. (2014) 312:1772–8. doi: 10.1001/jama.2014.13875

121. Gough E, Shaikh H, Manges AR. Systematic review of intestinal microbiota transplantation (Fecal bacteriotherapy) for recurrent *clostridium difficile* infection. *Clin Infect Dis*. (2011) 53:994–1002. doi: 10.1093/cid/cir632

122. Zipursky JS, Sidorsky TI, Freedman CA, Sidorsky MN, Kirkland KB. Patient attitudes toward the use of fecal microbiota transplantation in the treatment of recurrent *Clostridium difficile* infection. *Clin Infect Dis*. (2012) 55:1652–8. doi: 10.1093/cid/cis809

123. FDA. Commissioner O of the. Fecal Microbiota for Transplantation: Safety Alert - Risk of Serious Adverse Events Likely Due to Transmission of Pathogenic Organisms. (2020). Available at: https://www.fda.gov/safety/medical-product-safety-information/fecal-microbiota-transplantation-safety-alert-risk-serious-adverse-events-likely-due-transmission (Accessed March 7, 2023).

124. Zeng W, Shen J, Bo T, Peng L, Xu H, Nasser MI, et al. Cutting edge: probiotics and fecal microbiota transplantation in immunomodulation. *J Immunol Res*. (2019) 2019. doi: 10.1155/2019/1603758

125. Su S-H, Wu Y-F, Lin Q, Zhang L, Wang D-P, Hai J. Fecal microbiota transplantation and replenishment of short-chain fatty acids protect against chronic cerebral hypoperfusion-induced colonic dysfunction by regulating gut microbiota, differentiation of Th17 cells, and mitochondrial energy metabolism. *J Neuroinflamm*. (2022) 19:313. doi: 10.1186/s12974-022-02675-9

126. Davis BTI, Chen Z, Islam MBAR, Timken ME, Procissi D, Schwulst SJ. Postinjury fecal microbiome transplant decreases lesion size and neuroinflammation in traumatic brain injury. *Shock*. (2022) 58:287. doi: 10.1097/SHK.0000000000001979

127. Burrello C, Giuffrè MR, Macandog AD, Diaz-Basabe A, Cribiù FM, Lopez G, et al. Fecal microbiota transplantation controls murine chronic intestinal inflammation by modulating immune cell functions and gut microbiota composition. *Cells*. (2019) 8: E517. doi: 10.3390/cells8060517

128. Jacob V, Crawford C, Cohen-Mekelburg S, Viladomiu M, Putzel GG, Schneider Y, et al. Single delivery of high-diversity fecal microbiota preparation by colonoscopy is safe and effective in increasing microbial diversity in active ulcerative colitis. *Inflammatory Bowel Dis*. (2017) 23:903–11. doi: 10.1097/MIB.0000000000001132

129. Wang Y, Ren R, Sun G, Peng L, Tian Y, Yang Y. Pilot study of cytokine changes evaluation after fecal microbiota transplantation in patients with ulcerative colitis. *Int Immunopharmacol*. (2020) 85:106661. doi: 10.1016/j.intimp.2020.106661

130. Cenit MC, Sanz Y, Codoñer-Franch P. Influence of gut microbiota on neuropsychiatric disorders. *World J Gastroenterol*. (2017) 23:5486. doi: 10.3748/wjg.v23.i30.5486

131. Wang J-W, Kuo C-H, Kuo F-C, Wang Y-K, Hsu W-H, Yu F-J, et al. Fecal microbiota transplantation: Review and update. *J Formosan Med Assoc*. (2019) 118: S23–31. doi: 10.1016/j.jfma.2018.08.011

132. Chidambaram SB, Rathipriya AG, Mahalakshmi AM, Sharma S, Hediyal TA, Ray B, et al. The influence of gut dysbiosis in the pathogenesis and management of ischemic stroke. *Cells*. (2022) 11:1239. doi: 10.3390/cells11071239

133. Xie T, Yang R, Zhang X, Shen X, Yu L, Liao J, et al. Fecal microbiota transplantation alleviated cerebral ischemia reperfusion injury in obese rats. *Tohoku J Exp Med*. (2023) 259:49–55. doi: 10.1620/tjem.2022.J094

134. Spychala MS, Venna VR, Jandzinski M, Doran SJ, Durgan DJ, Ganesh BP, et al. Age-related changes in the gut microbiota influence systemic inflammation and stroke outcome. *Ann Neurol*. (2018) 84:23–36. doi: 10.1002/ana.25250

135. Wang J, Zhong Y, Zhu H, Mahgoub OK, Jian Z, Gu L, et al. Different gender-derived gut microbiota influence stroke outcomes by mitigating inflammation. *J Neuroinflamm*. (2022) 19:245. doi: 10.1186/s12974-022-02606-8

136. Wang H, Song W, Wu Q, Gao X, Li J, Tan C, et al. Fecal transplantation from db/db mice treated with sodium butyrate attenuates ischemic stroke injury. *Microbiol Spectr* (2021) 9:e00042–21. doi: 10.1128/Spectrum.00042-21

137. Chen R, Xu Y, Wu P, Zhou H, Lasanajak Y, Fang Y, et al. Transplantation of fecal microbiota rich in short chain fatty acids and butyric acid treat cerebral ischemic stroke by regulating gut microbiota. *Pharmacol Res*. (2019) 148:104403. doi: 10.1016/j.phrs.2019.104403

138. Lee J, d'Aigle J, Atadja L, Quaicoe V, Honarpisheh P, Ganesh BP, et al. Gut microbiota-derived short-chain fatty acids promote poststroke recovery in aged mice. *Circ Res*. (2020) 127:453–65. doi: 10.1161/CIRCRESAHA.119.316448

139. Du D, Tang W, Zhou C, Sun X, Wei Z, Zhong J, et al. Fecal microbiota transplantation is a promising method to restore gut microbiota dysbiosis and relieve neurological deficits after traumatic brain injury. *Oxid Med Cell Longev*. (2021) 2021:5816837. doi: 10.1155/2021/5816837

140. Wanchao S, Chen M, Zhiguo S, Futang X, Mengmeng S. Protective effect and mechanism of Lactobacillus on cerebral ischemia reperfusion injury in rats. *Braz J Med Biol Res*. (2018) 51:e7172. doi: 10.1590/1414-431x20187172

141. Wang Y, Tang J, Lv Q, Tan Y, Dong X, Liu H, et al. Establishment and resilience of transplanted gut microbiota in aged mice. *iScience*. (2022) 25:103654. doi: 10.1016/j.isci.2021.103654

142. Tan C, Wu Q, Wang H, Gao X, Xu R, Cui Z, et al. Dysbiosis of gut microbiota and short-chain fatty acids in acute ischemic stroke and the subsequent risk for poor functional outcomes. *J Parenteral Enteral Nutr*. (2021) 45:518–29. doi: 10.1002/jpen.1861

143. Xia G-H, You C, Gao X-X, Zeng X-L, Zhu J-J, Xu K-Y, et al. Stroke dysbiosis index (SDI) in gut microbiome are associated with brain injury and prognosis of stroke. *Front Neurol*. (2019) 10:397. doi: 10.3389/fneur.2019.00397

144. Stanley D, Moore RJ, Wong CHY. An insight into intestinal mucosal microbiota disruption after stroke. *Sci Rep*. (2018) 8:568. doi: 10.1038/s41598-017-18904-8

145. Yin J, Liao S-X, He Y, Wang S, Xia G-H, Liu F-T, et al. Dysbiosis of gut microbiota with reduced trimethylamine-N-oxide level in patients with large-artery atherosclerotic stroke or transient ischemic attack. *J Am Heart Assoc*. (2015) 4:e002699. doi: 10.1161/JAHA.115.002699

146. Chidambaram SB, Essa MM, Rathipriya AG, Bishir M, Ray B, Mahalakshmi AM, et al. Gut dysbiosis, defective autophagy and altered immune responses in neurodegenerative diseases: Tales of a vicious cycle. *Pharmacol Ther*. (2022) 231:107988. doi: 10.1016/j.pharmthera.2021.107988

147. Li N, Wang X, Sun C, Wu X, Lu M, Si Y, et al. Change of intestinal microbiota in cerebral ischemic stroke patients. *BMC Microbiol*. (2019) 19:191. doi: 10.1186/s12866-019-1552-1

148. Pleşeru A-M, Mihailă RG. The role of thrombin in central nervous system activity and stroke. *Clujul Med*. (2018) 91:368–71. doi: 10.15386/cjmed-973

149. Mohammed Y, Kootte RS, Kopatz WF, Borchers CH, Büller HR, Versteeg HH, et al. The intestinal microbiome potentially affects thrombin generation in human subjects. *J Thromb Haemostasis*. (2020) 18:642–50. doi: 10.1111/jth.14699

150. Zhuang Z-Q, Shen L-L, Li W-W, Fu X, Zeng F, Gui L, et al. Gut microbiota is altered in patients with alzheimer's disease. *J Alzheimers Dis*. (2018) 63:1337–46. doi: 10.3233/JAD-180176

151. Bonfili L, Cecarini V, Cuccioloni M, Angeletti M, Berardi S, Scarpona S, et al. SLAB51 probiotic formulation activates SIRT1 pathway promoting antioxidant and neuroprotective effects in an AD mouse model. *Mol Neurobiol*. (2018) 55:7987–8000. doi: 10.1007/s12035-018-0973-4

152. Dodiya HB, Kuntz T, Shaik SM, Baufeld C, Leibowitz J, Zhang X, et al. Sex-specific effects of microbiome perturbations on cerebral Aβ amyloidosis and microglia phenotypes. *J Exp Med*. (2019) 216:1542–60. doi: 10.1084/jem.20182386

153. Zhao Z, Ning J, Bao X, Shang M, Ma J, Li G, et al. Fecal microbiota transplantation protects rotenone-induced Parkinson's disease mice via suppressing inflammation mediated by the lipopolysaccharide-TLR4 signaling pathway through the microbiota-gut-brain axis. *Microbiome*. (2021) 9:226. doi: 10.1186/s40168-021-01107-9

154. Gorecki AM, Anyaegbu CC, Anderton RS. TLR2 and TLR4 in Parkinson's disease pathogenesis: the environment takes a toll on the gut. *Transl Neurodegener*. (2021) 10:47. doi: 10.1186/s40035-021-00271-0

155. Li Y, Xia Y, Yin S, Wan F, Hu J, Kou L, et al. Targeting microglial α-synuclein/TLRs/NF-kappaB/NLRP3 inflammasome axis in parkinson's disease. *Front Immunol*. (2021) 12:719807. doi: 10.3389/fimmu.2021.719807

156. Scheperjans F, Derkinderen P, Borghammer P. The gut and parkinson's disease: hype or hope? *J Parkinsons Dis*. (2018) 8:S31–9. doi: 10.3233/JPD-181477

157. Huang H, Xu H, Luo Q, He J, Li M, Chen H, et al. Fecal microbiota transplantation to treat Parkinson's disease with constipation: A case report. *Med (Baltimore)*. (2019) 98:e16163. doi: 10.1097/MD.0000000000016163

158. Zhong Z, Chen W, Gao H, Che N, Xu M, Yang L, et al. Fecal microbiota transplantation exerts a protective role in MPTP-induced parkinson's disease via the TLR4/PI3K/AKT/NF-κB pathway stimulated by α-synuclein. *Neurochem Res*. (2021) 46:3050–8. doi: 10.1007/s11064-021-03411-0

159. Matta SM, Hill-Yardin EL, Crack PJ. The influence of neuroinflammation in Autism Spectrum Disorder. *Brain Behav Immun*. (2019) 79:75–90. doi: 10.1016/j.bbi.2019.04.037

160. Inga Jácome MC, Morales Chacòn LM, Vera Cuesta H, Maragoto Rizo C, Whilby Santiesteban M, Ramos Hernandez L, et al. Peripheral inflammatory markers contributing to comorbidities in autism. *Behav Sci*. (2016) 6:29. doi: 10.3390/bs6040029

161. Hughes HK, Rose D, Ashwood P. The gut microbiota and dysbiosis in autism spectrum disorders. *Curr Neurol Neurosci Rep*. (2018) 18:81. doi: 10.1007/s11910-018-0887-6

162. Singh RK, Chang H-W, Yan D, Lee KM, Ucmak D, Wong K, et al. Influence of diet on the gut microbiome and implications for human health. *J Trans Med*. (2017) 15:73. doi: 10.1186/s12967-017-1175-y

163. Eshraghi RS, Davies C, Iyengar R, Perez L, Mittal R, Eshraghi AA. Gut-induced inflammation during development may compromise the blood-brain barrier and predispose to autism spectrum disorder. *J Clin Med*. (2020) 10:27. doi: 10.3390/jcm10010027

164. Fattorusso A, Di Genova L, Dell'Isola GB, Mencaroni E, Esposito S. Autism spectrum disorders and the gut microbiota. *Nutrients*. (2019) 11:E521. doi: 10.3390/nu11030521

165. Jain M, Swaminathan S, Teresa S, Srividhya G, Govindarao B, Gopal L, et al. Fecal microbiota transplantation for autism. *Trop Gastroenterol*. (2021) 41:147–8. doi: 10.7869/tg.599

166. Chen K, Fu Y, Wang Y, Liao L, Xu H, Zhang A, et al. Therapeutic effects of the *in vitro* cultured human gut microbiota as transplants on altering gut microbiota and improving symptoms associated with autism spectrum disorder. *Microb Ecol*. (2020) 80:475–86. doi: 10.1007/s00248-020-01494-w

167. Goo N, Bae HJ, Park K, Kim J, Jeong Y, Cai M, et al. The effect of fecal microbiota transplantation on autistic-like behaviors in Fmr1 KO mice. *Life Sci*. (2020) 262:118497. doi: 10.1016/j.lfs.2020.118497

Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Quantum Image processing algorithms comparison using datasets of machine learning based applications

BARKHA SINGH (singh.barkha718@gmail.com)
Delhi Technological University
SreeDevi Indu
Delhi Technological University
Sudipta majumdar
Delhi Technological University

Research Article

Keywords: Twin Support Vector Machines (TWSVM), Quantum enhanced Support vector machine(QeSVM), Dynamic Quantum clustering (DQC), Quantum

Posted Date: February 12th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3939791/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Signal,image and Video processing manuscript No.
(will be inserted by the editor)

Quantum Image processing algorithms comparison using datasets of machine learning based applications

Barkha Singh · Sreedevi Indu · Sudipta Majumdar

Received: date / Accepted: date

Abstract This article analyzes and compares various Quantum learning algorithms on big data. The main contribution of this article is to provide a new machine-learning approach using Quantum computing for big data analysis with features of robust, novel, and effective Quantum machine images and transit mode. Currently, there is no efficient method for doing Quantum image recognition or classification due to the lack of an effective Quantum feature extraction technique. This work proposes a global Quantum feature extraction technique based on Schmidt decomposition for the first time. Additionally, a new version of the Quantum learning algorithm is shown, which uses the features' Hamming distance to classify images. With the help of algorithm analysis and experimental findings from the benchmark database Caltech 101, a successful method for large-scale image classification is developed and put forth in the context of big data. The proposed model yields an with an average accuracy of 98% with the proposed enhanced Quantum classifier, QeSVM classifier, swarm particle optimizer with Twin wave SVM, QPSO-TWSVM, and other Q-CNN models on different Big Data sets.

Keywords Twin Support Vector Machines (TWSVM) · Quantum enhanced Support vector machine(QeSVM) · Dynamic Quantum clustering (DQC) · Quantum Variational Support Vector Machine (QV-SVM) · Quantum Particle Swarm Optimization (QPSO) · Generalized eigenvalue proximal SVM (GePSVM)

Barkha singh* (corresponding author)
Dept. of ECE, DTU, Delhi
E-mail: singh.barkha718@gmail.com

S. Indu
Dept. of ECE, DTU, Delhi
E-mail: s.indu@dce.ac.in

S. Majumdar
Dept. of ECE, DTU, Delhi
E-mail: sudiptamajumdar@dtu.ac.in

1 Introduction

Machine learning models predict the future by making decisions according to previous pattern analysis. Big data plays its role when a large amount is collected from multiple sources and further processed to analyze the relationship and pattern for various applications. Big data, as the name describes, is "big" data in amount, diverse in nature, and has structured and Unstructured data types that characterize data by its volume, variety, variability, and value. This article discusses big data analysis using Quantum Machine learning. Because of the basic unit (photons), the heterogeneity of big data could be optimized. Quantum computing for big data analysis will be a scorching research area. Different machine learning techniques have their own Quantum versions in the form of Quantum classification, Quantum clustering, and Quantum searching for areas such as computer science, Bioinformatics, financial analysis [1], and robotics-based automation. As concluded in this paper, big data has achieved an exponential growth rate, which comes with its computational challenges that can be solved better by Quantum machine learning approaches [2].The manipulation of matrices and vectors using Quantum computing helps achieve an advantageous exponential speed over classical as per experimental analysis [3][4].For critical feature extraction and for multi-category image classification data, Quantum CNN model is used by proposing a hybrid Quantum-classical convolution neural network (Q-CNN) for both classical and Quantum parts of image processing [5].

The efficient amplitude encoding technique is also analyzed on many datasets for image classification using the Tensor Flow Quantum platform [6] as shown in figure 9. It has been analyzed that QC-CNN manifests the generalizability due to Quantum parallelism ,a Quantum property that sped up Quantum state transformation simultaneously.

2 Impact of Quantum Computing on Big Data Analysis

Quantum machine learning helps in finding and assigning clusters using supervised and un-supervised Quantum machine learning algorithms [7]. Quantum machine learning applications for big data analysis are relatively challenging. The proposed work in this field could be considered an initiative for future research on big data analysis. The solution to complex problems is quite unachievable because of massive data sets [8]. A Quantum computer-based approach reduces the inevitable distortion of "bad data" arising during the collection of large amounts of data from the real world. The Quantum-based approach is used to exponentially speed up complex calculations and help in obtaining all significant features of the data [9]. Also, for reconstructing the pattern, despite changes such as stretching, compression, or distortion during processing.If a data set has 'N' points, the conventional "Big data" approach would need '2N' processing units, whereas the Quantum computing based approach would need only 'N' processing units. Therefore, this approach is also helpful for device size reduction . This ground-breaking (Quantum-based) technology can unlock the hidden information(pattern) of a massive data set [10]. In addition, the superposition of state is achieved in Quantum computing. Hence, certain types of calculations can be performed much faster than those performed by classical [11]. By harnessing the power of parallelism and reducing computational complexity, Quantum computing helps in Big data analysis. By utilizing the properties of Quantum mechanics, Quantum computing can process vast amount of data simultaneously [12]. Quantum computing offers a promising solution by providing the capability to handle and process massive data sets in a fraction of the time compared to classical computers. The ability to leverage Quantum computing in big data analysis can lead to groundbreaking insights and advancements across various domains [13] [14]. It is important to note that the practical implementation of Quantum computing for big data analysis is still in its early stages. The development of Quantum computational methods allows for their use in areas such as machine learning, decision models, and big data analysis. These advanced computational methods offer the potential for significant advancements in data analysis, enabling more accurate predictions, deeper insights, and efficient decision-making. The use of Quantum computing in big data analysis has the potential to revolutionize the field by allowing for faster and more efficient processing of massive data sets [15]. This can lead to groundbreaking insights and advancements across various domains, enabling more accurate predictions, deeper insights, and efficient decision-making. The ability to process vast amounts of data simultaneously through Quantum computing can provide a competitive advantage in machine learning applications and enable the healthcare industry [16] [17] to evaluate and treat complicated medical disorders more effectively. Overall, Quantum computing has the potential to revolutionize big data analysis by enabling the processing of exponential amounts of data in a fraction of the time compared to classical computers. The combination of Quantum computing and medical systems provides a sophisticated solution that allows for more accurate evaluation and treatment of complicated medical disorders [18] [5]. In summary, the development of Quantum computational methods allows for their use in areas such as machine learning, decision models, and big data analysis [19]. The use of Quantum computing in big data analysis offers the potential for significant advancements in data analysis, enabling more accurate predictions. Faster and more efficient processing capabilities can lead to breakthroughs in various fields such as internet security, pharmacology [18], complex problem-solving, and Artificial intelligence [20]. The use of Quantum computing in big data analysis is a promising avenue for research and development. It holds the key to unlocking new possibilities in data analysis [21], providing researchers and industries with the tools to gain a competitive edge and make groundbreaking discoveries. With the advantage of fast decision-making and speeding up complex calculation solutions, Quantum computing can solve some problems of big data analytics [12] such as optimization, matrix inversion, and factorization. From identifying patterns and analyzing trends and correlations, it is difficult to achieve all this with a good optimization level [22].

Certain classical encryption methods are unable to secure large amounts of sensitive data therefore, Quantum-resistant encryption techniques could play an advantageous role [23]. Quantum image processing yields multi-particle Quantum states that are advantageous for parallelism. This section lists all the existing Quantum datasets in Section 3 and Miscellaneous datasets have been mentioned in Section 4 and 5.

3 Development and validation of Quantum models based Datasets

Speeding up complex calculations, better pattern recognition, fast database searching, query optimization, data compression, dimensionality reduction, cryptography, and data security at the optimum level are some of the advantages of Quantum techniques based on big data analysis.

3.1 Quantum Machine Learning Algorithms for Big Data Analysis

There are two main classes of Quantum algorithms, Grover's algorithm and Shor's Algorithm.Shor's algorithm is used generally for factoring and Grover's algorithm is used basically for the unstructured search [24]. Shor's algorithm is a quantum computer algorithm used for integer factorization. This is a polynomial-time quantum computer whereas, Grover's Algorithm is used for Searching from an unsorted database that can be sped up within data sets by this Quantum search algorithm that helps to achieve a Quadratic speedup. This algorithm searches the uniqueness with high probability from a black box function having domain size of 'N', O($\sqrt{N}$) number of function calls unlike classical algorithms that take O(N) function calls.

– Mathematical steps for Grover's algorithm are defined and designed by following steps . In this, 'N' items in an unsorted database have a fixed target item whose index is selected by applying Grover's algorithm.[25]

Quantum step preparation It is the preparation of a uniform superposition that consists of all possible states.

Oracle An operator, represented by 'U' that represents the 'target' amplitude near the Quantum state and leaves 'non-target amplitude' untouched.

Amplitude amplification (Diffusion operator) This step involves two reflections one is, the mean and one is, $|s\rangle$ axis [26].

Algorithm iteration Algorithm iteration is also known as 'optimal no. of iterations' is an Amplitude amplification that steps around $\sqrt{N}$ times for increasing the correct index measuring probability.

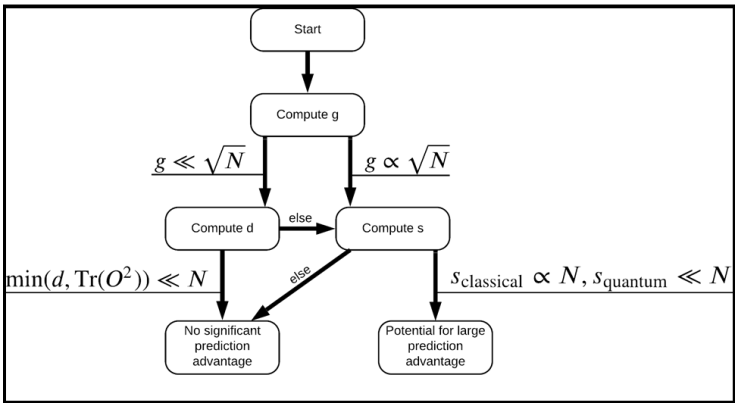


Fig. 1 Algorithm for Quantum Data

Measurement This step collapses iteration to one of the computational basis states by measuring the Quantum state. For example, for N–dimensional vectors and M representations samples in one of the K-clusters , o(log(MN)) is the total time taken by Quantum computing which contributes to the potential wide application where convergence takes O(poly(MN)). For linear equation system solutions, many Quantum algorithm and simulations like Shor's algorithm provides a great role described by Shor's [27].

3.2 Health care data of cardiovascular events

An experimental analysis of article with the development and validation of QRISK3 [13] for risk prediction to estimate future risk of cardiovascular disease with Health care data of cardiovascular events in men and women for the meta-analysis of C-statistics is done in figure 2.

3.3 Quantum PSO Algorithm for good fitness analysis Data

QPSO is an optimization algorithm that combines principles of quantum computing and particle swarm optimization. Particle swarm optimization is a nature-inspired optimization technique where a population of particles evolves in a search space to find the optimal solution to a problem [28]. In QPSO, quantum principles are incorporated to enhance the optimization process. Quantum-inspired operators are used to update the particle positions and velocities, potentially lead to more efficient optimization than classical particle swarm optimization.with the help of fitness dataset,existed PSO algorithm is restructured with Quantum properties as shown in figure 3 [29]. The QPSO algorithm

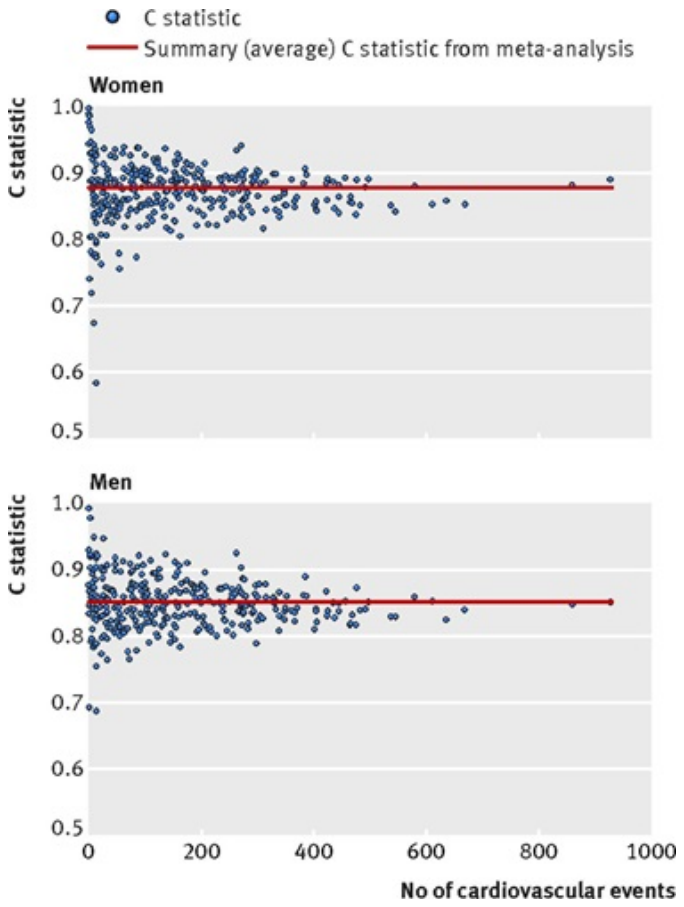


Fig. 2 cardiovascular events in men and women for the Quantum meta-analysis

Algorithm 1 Quantum particle swarm optimization (QPSO)

Initialize the particles' positions ($\mathbf{X}^{(t)}(i)$) and the number of particles N.
Set $\mathbf{P}^{(t)}(i) = \mathbf{X}^{(t)}(i)$, $\forall\, i = 1, 2, \ldots, N$.
while ($t < Max_{iter}$) **do**
  for all $i = 1$ to N **do**
    Calculate $\mathbf{M}_{best}$ as in Eq. 13.
    Evaluate the fitness function for the current position ($\mathcal{F}(X^{(t)}(i))$).
    if ($\mathcal{F}(\mathbf{X}^{(t)}(i)) < \mathcal{F}(\mathbf{P}^{(t)}(i))$) **then**
      $\mathbf{P}^{(t)}(i) = \mathbf{X}^{(t)}(i)$.
    end if
    $\mathbf{P}^{(t)}(g) = min(\mathcal{F}(\mathbf{P}^{(t)}(i)))$.
    for $j = 1$ to d **do**
      $\phi =$ rand(0,1).
      $p(i, j) = \phi P^{(t)}(i, j) + (1 - \phi) P^{(t)}(g, j)$.
      $u =$ rand(0,1).
      $k =$ rand(0,1).
      if ($k \geq 0.5$) **then**
        $X^{(t)}(i, j) = p(i, j) - \beta.|(M_{best}(i, j) - X^{(t)}(i, j)|.ln(1/u)$.
      else
        $X^{(t)}(i, j) = p(i, j) + \beta.|(M_{best}(i, j) - X^{(t)}(i, j)|.ln(1/u)$.
      end if
    end for
  end for
  Stop the algorithm if a sufficiently good fitness function is met.
end while

Fig. 3 Quantum particle swarm optimization for fitness dataset

is one of the optimization algorithms in which by initializing the particle position with the 'ith' number of X_(t)(i) for all 'i' started from 0 to 'N'.M_(best) as per equation needs to be calculated for all values of 'i' and the current position , fitness function is evaluated for (f(X(t)_(i))).In the loop ,if (f(x_(t)(i)) will be less than f(p_(t)(i)) then the value of function $\hat{p(t)}(i)$).for the 'j' value upto dimension 'd' the 'ϕ' value will be equal to random value between 0 to 1 and the function p(i,j) will be equal to the sum of $\phi(p_(t)(i,j), (1-\phi)P_(t)(g,j))$ and u and k value will be equal to random values of 0 and 1.for the k greater than equal to 0.5 the function $\hat{X(t)}(i,j)$ is the subtraction of z equal to the β | $(M_best(i,j) - \hat{X(t)}(i,j)$ | .ln(1/u) from the P(i,j) function otherwise if value of K will be less than 0.5 the 'Z' will get added with function P(i,j) and the algorithm get summarized and got ended if the optimized fitness function met.

3.3.1 Quantum-Inspired PSO (QIPSO)

Classical PSO gets combined with a Quantum-inspired mechanism for making search algorithms optimized and good exploration.

3.3.2 Hybrid Quantum PSO (HQPSO)

For a subset of particles, the Quantum-inspired mechanism is used and Classical PSO is used for the remaining particles. This hybridization is done by HQPSO algorithm.

3.3.3 Quantum-Behaved PSO (QPSO-II)

For better exploration and convergence rates, the Quantum behavior of particles should be dynamically adjusted for better optimization, a strategy introduced by QPSO-II [15].

3.3.4 Multi-Objective QPSO (MO-QPSO)

By MO-QPSO ,Pareto optimal solutions for multi-objective optimization problems, A multi-objective optimization framework is made inside Quantum particles.

3.3.5 Quantum Annealing PSO (QAPSO)

For the global optimum and better search process Quantum annealing is considered along with PSO for better optimization.

3.3.6 Quantum Entanglement in QPSO

Good optimization performance is obtainable by entanglement using QPSO. Basically, for the distribution of information in many complex ways Quantum entanglement plays an important role [30].

3.3.7 Adaptive QPSO (AQPSO)

By adjusting all the Quantum particle's behavior and parameters with some adaptive strategies, AQPSO solves different problems.

3.3.8 Parallel QPSO (PQPSO)

The algorithm convergence speed up by the parallelism concept of PQPSO that optimizes multiple particles simultaneously.

3.3.9 Quantum-Inspired Differential Evolution with PSO (QIDPSO)

For all the optimization problems QIDPSO and PSO get combined for differential evolution which is Quantum inspired [29].

3.4 Quantum techniques utilization for Big data analysis [20]

Speeding up complex calculations, better pattern recognition, fast database searching, query optimization, data compression, dimensionality reduction, cryptography, and data security at the optimum level are some of the advantages of Quantum techniques based on big data analysis.Many documents published in previous years to showcase the same as shown in figure 4.[27].

For example, for N–dimensional vectors and M representations samples in one of the K-clusters , o(log(MN)) is the total time taken by Quantum computing which contributes to the potential wide application where convergence takes O(poly(MN)). For linear equation system solutions, many Quantum algorithm and simulations like Shor's algorithm provides a great role described by Shor's algorithm.

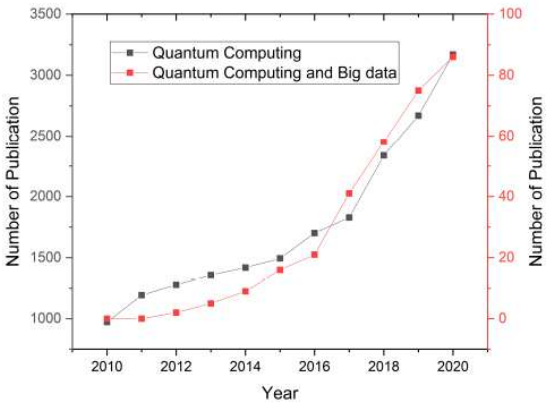


Fig. 4 No. of Quantum computing and Big data image analysis year wise published work

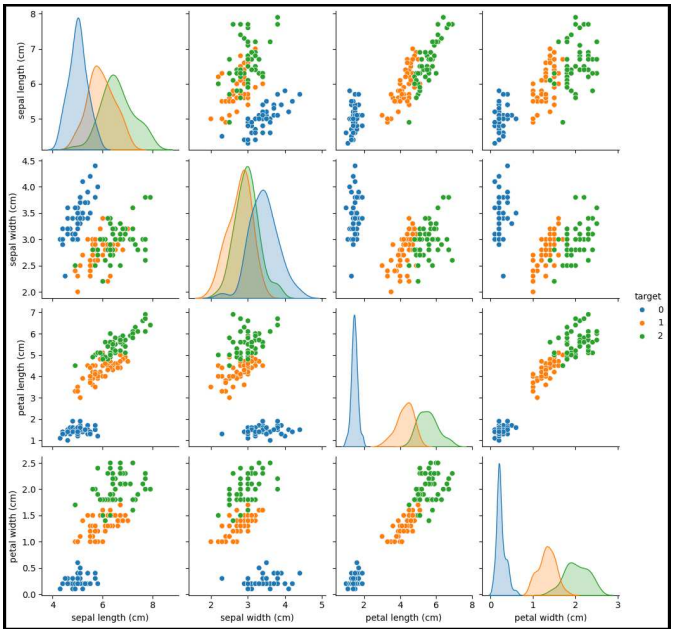


Fig. 5 QSVM modal on IRIS Dataset

3.5 Quantum Image classifiers based analysis

3.5.1 Quantum Support Vector Machines (QSVM) on IRIS DATA SET

QSVM is a Quantum version of SVM classifies which set the data points as per Quantum computing principle into different categories shown in figure 5 . [31]. For the data point classification,QSVM uses quantum computing methods described in the articles[32] [33].Using IRIS_DATA SET , QSVM results has showcased in figure 6.

```
============== ==== ==== ======= ===== ====================
                Min  Max   Mean    SD   Class Correlation
============== ==== ==== ======= ===== ====================
sepal length:   4.3  7.9   5.84   0.83    0.7826
sepal width:    2.0  4.4   3.05   0.43   -0.4194
petal length:   1.0  6.9   3.76   1.76    0.9490  (high!)
petal width:    0.1  2.5   1.20   0.76    0.9565  (high!)
============== ==== ==== ======= ===== ====================

:Missing Attribute Values: None
:Class Distribution: 33.3% for each of 3 classes.
```

Fig. 6 Different categories of IRIS Dataset

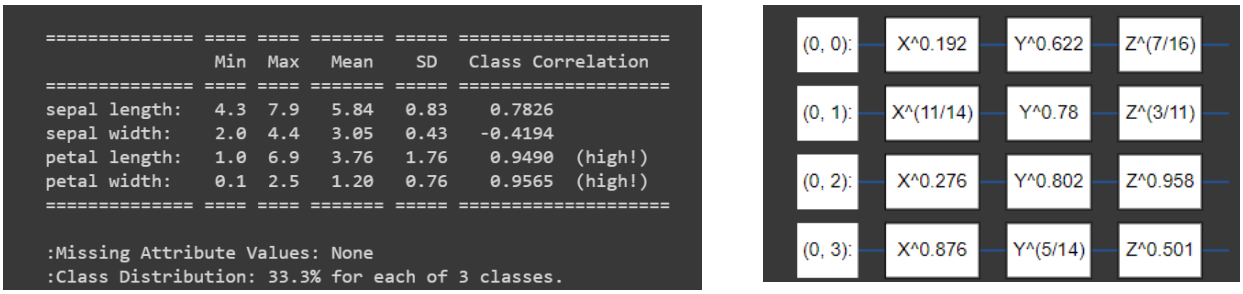


Fig. 8 Quantum VSVM Classifier circuit

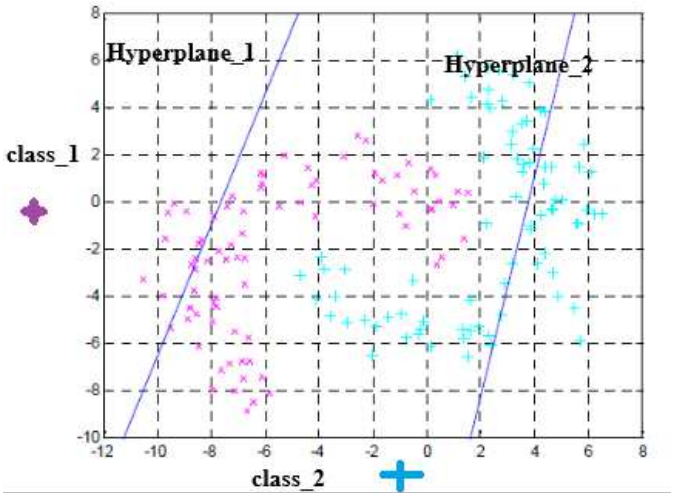


Fig. 7 Twin SVM classifier of two classes

3.5.2 Twin Support Vector Machines (TWSVM)

Twin Support Vector Machines are made for binary classification tasks. Unlike traditional SVMs, which aim to find a single hyperplane that separates data into two classes, TWSVM seeks to find two non-parallel hyper planes that maximize the margin between the data points of the two classes. These hyperplanes are called "twins" shown in figure 7. TWSVM is particularly useful when dealing with imbalanced data sets, as it can provide a better trade-off between classification accuracy and can handle the class imbalance issue [7].

3.5.3 Quantum Variational Algorithms

By optimizing the parameters of the Quantum circuit [34], Supervised machine learning tasks could be performed by Variational quantum algorithms, such as the Quantum Variational Support Vector Machine (QVSVM) and the Quantum Variational Classifier (QVC) as shown in figure 10 that describes the objective function value and iteration chart.[35] .

3.5.4 Quantum Neural Networks (QNNs)

Quantum versions of classical neural networks are called Quantum neural networks used for classification and regression tasks.For unsupervised machine learning tasks, Quantum variants of k-means clustering algorithms are being explored, such as data clustering [36] .Quantum Principal Component Analysis (PCA) For speeding up high dimensional data processing, Quantum PCA algorithms are used efficiently."Variational Quantum Optimization of Non-locality in Noisy Quantum Networks [35]" discussed the IBM Quantum network [35] and Variational Quantum optimization (VQO) framework in which as per the number of Qubits, Quantum Noise increases [8] and shows that on IBM Quantum computer, VQO technique can optimize non-n-locality. For Noise intermediate scale Quantum (NISQ) devices, Hybrid Quantum-classical algorithm is also proposed [37]. In the experimental analysis, PennyLane, an open-source and free Quantum differential programming framework is used, and Quantum network variational optimizer qNetVO using a Python package is shown in figure 8 [33]. For probabilistic graphical modeling and unsupervised learning, Quantum versions of classical Boltzmann Machines is considered. [38].

3.5.5 Twin support vector machines based on fruit fly optimization algorithm

SVM which is also known as an optimized binary classifier, is a supervised machine learning algorithm used for new data classification which is an example of already classified training [33]. With the logarithmic complexity having a size of vectors, SVM can be implemented using a Quantum computer. QSVM works on logarithmic run time for a full training set unlike classical sampling theorems which require polynomial time, the Quantum environment gives an exponential speed up [13] . A supervised Machine learning algorithm takes the training data from the sets and classifies vectors in a feature space into one of two states [39].SVM constructs the optimal hyperplane that divides the two sets

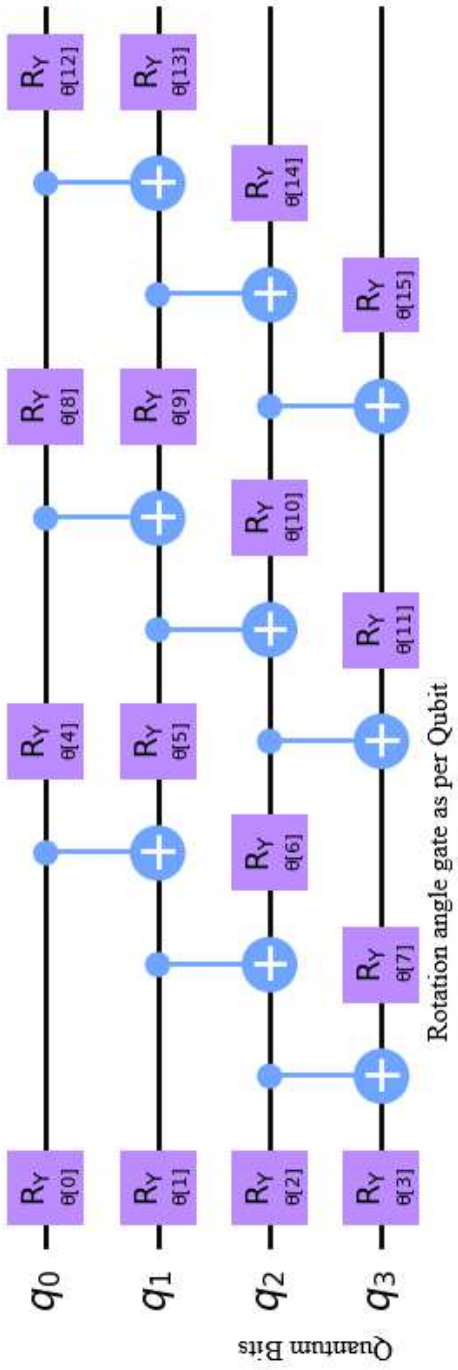


Fig. 9 Variational Quantum optimization framework

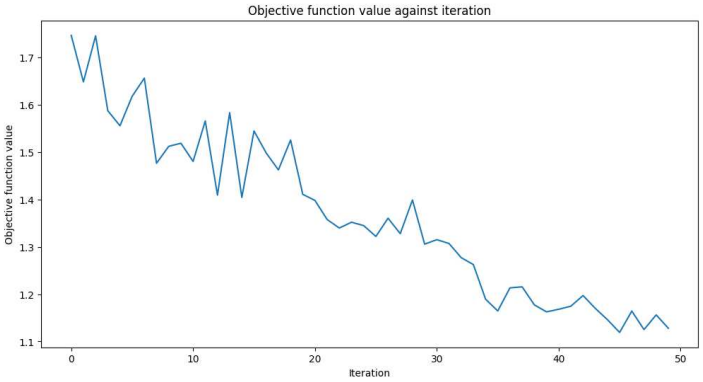


Fig. 10 Variational Quantum computing VQC objective function value vs iteration chart

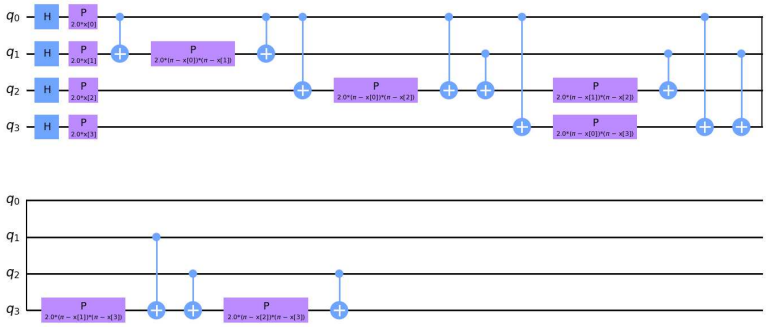


Fig. 11 A parameterized Quantum circuit, layers of a neural network, with tunable parameters or weights that minimize the objective function

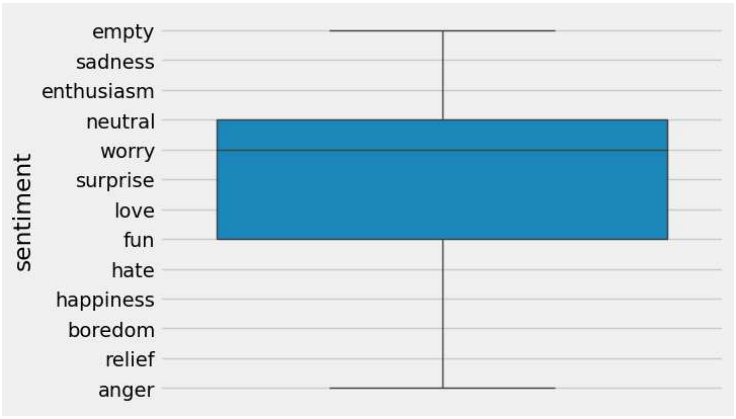


Fig. 12 Real-Time sentiment analysis Box plot using QSVM

into a high-dimensional kernel space and an original feature space.SVM is considered a Quadratic programming problem and also SVM solves time proportional to $O(log(\epsilon^{(-1)})poly(N, M)$ where 'N' is the dimensional of feature space and 'M' is several training vectors and ϵ is the accuracy [6]. At both the training and classification stage QSVM was implemented with 0(log(NM)) run time. As per the inner products, the performance of 'N' increases for better Quantum evaluation. whereas, For better performance of 'M', the Matrix inversion algorithm is used that solves the least square problem. In the Quantum environment, QRAM is considered to use O(MN) hardware resources and access them by O(log MN) For performing a matrix inversion of training data efficiently, a non-sparse matrix exponential training is used under the Big data algorithm. According to the problem and complexity, the Quantum algorithm can

have exponential or polynomial speed compared to the classical algorithm [28]. Machine learning has supervised and unsupervised learning, for high dimensional vectors and corresponding labels for known and unknown categorizing sets of training examples respectively. [40]. Since, classification is a process of supervised learning .A new Quantum pattern classification algorithm introduced with Quantum machine learning for hamming distance calculation in Quantum computer environment .Algorithm that performs Linear regression with least square optimization is also discussed in [41].

[32] proposed a weighted K-nearest neighbor method for pattern classification on a Quantum computer in which author explained if the input is Quantum information, with 'N' dimension of feature vectors and training set size is independent then run time order will be the order of logarithmic of O(log N) instead of 'X' training inputs matrix. since, Quantum based algorithm merely requires 'x*x' represented by low-rank approximation.

[23] introduced a Quantum version of the Discrete-Time classifier to determine nodes that are to be split for Quantum criterion.Then, the measurement between two Quantum states is used to subclass the training data for clustering by considering two different data sets, and a comparison of Quantum computing performance with other conventional classical classification methods (SVM, K-NN) is done by [4].

[20] explained the Q-SVM for big data classification to implement both linear and non-linear exponentially to speed up Quantum Binary classifier in its optimized form in which the size of vectors and no. of training examples are large. The author also discussed a non-sparse matrix simulation technique along with inversion of the matrix for efficient performance of PCA, used for data training.

[6] introduced Generalized Eigenvalue proximal SVM and proposed DSA-GEPSVM (Data structure algorithm-GEPSVM) to solve complexity in working with error or noise effect data sets. [10] discusses the high computational cost of K-SVM for large data sets. For use in Big data analytics, the Author proposed a divide-and-conquer mechanism for training K-SVM. A Buffered K-D Tree is used for dividing the data samples. A buffered K-D Tree also known as a Balanced Binary tree is used to divide and conquer data samples into many sets also known as leaves. Each cell Buffer is used as a single vector, and a strategy named 'min-max' expansion strategy is proposed as per data samples buffered in the K-D Tree, The KSVM Classifier is updated according to the tree [27]. For effective results, two Benchmark data sets are compared with two state-of-the-art training methods, and Lib SVM is used for KSVM training.

[24] discussed the local multiple kernel learning algorithm, a soft clustering and classification method. To address the issues discussed above [27], a soft clustering-based local multiple shows that the kernel weights solved by the algorithm are better suited to the characteristics of the data sets.LMKL is a sample-based proposed method and all experiments are done on the UCI data set. The author compared the results of SVM with a single Gaussian kernel [41]. [6] 2023 presented ideas from Quantum physics to help and create a machine learning application using Quantum computing specifically for optimizing the computational complexity of massive data sets calculations [42]. Quantum Big Bang, a creature of Quantum computers, its assessment, and development are well explained by the author [1], they also discussed Quantum parallelism with its superposition and linearity concept that is helpful for infinite no. of inputs.DiVincenzo criteria are used in Quantum computing technologies, complemented with its specific physical layer feature [43]. The author also discusses the application of Big data analysis in the area of personal health care scientific discoveries to social interaction, information is inundated in every element of our lives so the generation of data will keep increasing enormously and unwieldy. Hence, there is a requirement for large-scale Big data analysis [18], the author discusses privacy as a big challenge for Big data analysis. [34] in their article discussed the hybrid of PSVM (proximal-SVM) and GEPSVM (Proximal SVM based on the generalized eigenvalues) and proposed TWSVM model. A Big-binary classification programming problem is converted into two smaller Quadratic programming by TWSVM by getting two Non-parallel hyperplanes. Problems faced in the above approach are solved by the Swarm optimization approach named QPSO-TWSVM which is Quantum particle swarm optimization TWSVM. since QPSO has its global search ability that can help in searching the optimal parameters. Instead of selecting the parameters blindly, this proposed approach appoints parameters accurately in comparison to traditional TWSVM.Ten data sets are used to showcase linear test results with no. of samples and attributes is shown in figure 13 and corrosponding Quantum classifier modal accuracy comparison is shown in figure 14.

Least Squares SVM was discussed by Suykens et al [28]. Proximal SVM is used for setting a hyperplane in each type of sample point to solve all Binary classification problems by making two parallel hyperplanes and optimizing the distance between two parallel hyperplanes by using the equality constraints and can-

Data sets	Samples	Attributes
Australian	690	14
Sonar	208	60
Pima-Indian	768	8
Ionosphere	351	34
BUPA liver	345	7
Heart-statlog	270	14
CMC	1473	9
WPBC	198	34
Hepatitis	155	19
Heart-c	303	13

Fig. 13 Ten datasets description for linear test

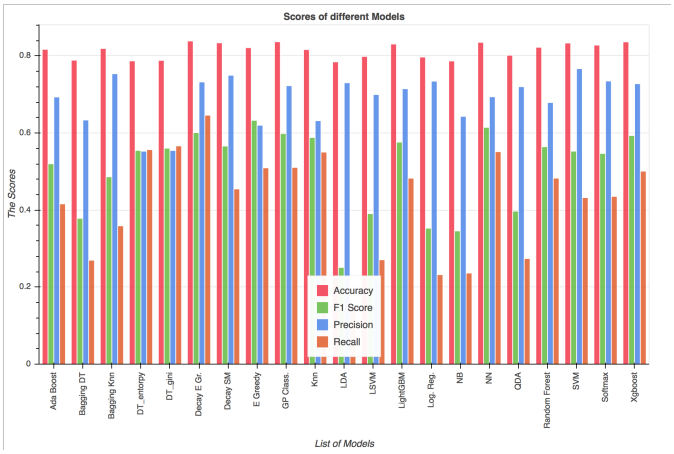


Fig. 14 Quantum classifier modals accuracy comparison

celing the inequality constraints which is the character of traditional SVM.GEPSVM is introduced that cancels constraints to make two hyperplanes in parallel in PSVM. One Eigenvalue problem is converted into two generalized smallest eigenvalue problems for obtaining global extreme [39].

[14] introduced the algorithm of TWSVM by deep studying PSVM and GEPSVM in 2007 in which two hyperplanes have no constraints of parallel condition.

[18] introduced WLSTWSVM, weighted least square TWSVM whereas [39] introduced robust TWSVM for better pattern classification in 2009. since TWSVM showed good classifier, results for real-time processing.Sentiment Analyis using real time sentiment based data using Quantum classifier[8] is shown in figure 12.

Duan Jingbo [29] in their article discussed "Big Data Classification Model" and algorithm based on "Double Quantum Particle Swarm Optimization" considered the use of intelligent algorithms to model identification of multi-variable systems has become a mainstream research field for traditional large data classification models and algorithms, which have the disadvantage of extended processing time. This research suggests a practical technique for data-parallel optimized computing. The difficulties in identifying multi-variable systems that merge historical big data and intelligent algorithms are investigated, and the impact of input variables on output is assessed. By Creating a big data classification model, valid data points, arranging data combination methods, dividing data local connection methods, obtaining local node micro-cluster data, calculating redundant data and invalid data increment values in node data groups, and reconstructing central node sample algorithm [44]. By Creating the experiment, replicating the testing setting, and confirming that the suggested model calculation technique can reduce the amount of time needed to process data in real-world applications. This has research implications for practice. It can significantly reduce communication costs between network nodes, improve global mining accuracy, and improve query efficiency in distributed data. It also has good utilization value in the analysis of massive data information. The algorithm is designed with data classification to form a new data set, and then its steps are described briefly. Finally, simulation experiments are used to verify the algorithm's viability.

[39] Quantum mechanics can speed up a range of search applications over unsorted data. For example, imagine a phone directory containing 'N' names arranged in completely random order. To find someone's phone number with a probability of 50 percent, any classical algorithm (whether deterministic or probabilistic) will need to access the database a minimum of '0.5N' times. Quantum mechanical systems can be in a superposition of states and simultaneously examine multiple names. By properly adjusting the phases of various operations, successful computations reinforce each other while others interfere randomly. As a result, the desired phone number can be obtained with only $O(\sqrt{N})$ access to the database.

[45] in their article "Simultaneous execution of Quantum circuits on current and near-future NISQ system" explained the parallel allocation of Quantum computer also known as "Palloq" " a compilation protocol that optimizes the Quantum multi-programming performance on NISQ processor. In their experimental analysis, the Authors focused on the No. of Quantum computers executed by using Quantum multi-programming, total execution time improves the throughput, and by executing multiple Quantum computers simultaneously, the NISQ processor is utilized at the optimized level. [39] uses the DQPSO Algorithm to improve MBSVM Parameters that focus on CE (Contract Expansion) coefficient control mode to realize the Dynamic regulation of CE coefficient α. Better optimization performance and fast convergence speed of the DQPSO algorithm as compared to classical QPSO is explained by [29]. The proposed DQPSO-mbsvm algorithm is used for MBSVM classification performance analyzed in [21]. [6] in their article entitled Amplitude Embedding which is an ex-

Table 1 List of acronyms

Acronym	Definition
DFT	Discrete Fourier Transform
QFT	Quantum Fourier Transform
QSVM	Quantum Support Vector
CSV	Comma Separated Values
QeSVM	Quantum enhanced Support Vector
DT-CWT	Dual-tree complex wavelet transform
FFT	fast Fourier transform
TWSVM	Twin Wave Support Vector
QPSO	Quantum Particle Swarm Optimization
DWT	Discrete wavelet transform
DQC	Dynamic Quantum clustering
QVSVM	Quantum Variational Support Vector
GEPSVM	Generalized eigenvalue proximal SVM
QCNN	Quantum-classical convolution neural
PCA	Principal component analysis
DQC	Dynamic Quantum clustering
VQO	Variational Quantum optimization
NISQ	Noise intermediate scale Quantum
QNVO	Quantum network variational optimizer
HQPSO	Hybrid Quantum PSO
DWP	Discrete wavelet packets
SVM	Support vector machine
K-NN	K Nearest neighbours
RNN	Recurrent neural network
QRAM	Quantum Random Access Memory
CNN	Convolutional neural network
LS-SVM	Least-squares support-vector machine

ponential scaling. In this embedded technique,'n' qubits consist of '2n' classical data values. Since, Qubits, a basic unit that has information consisting of $\mid 0 \succ$ and $\mid 1 \succ$ basis states by superposition principle in a 2-D Quantum system and Qudits for d-dimensional Quantum system in Quantum computing $\mid 0 \succ$, $\mid 1 \succ$ up to $\mid (d) \succ$ are the 'd' basis state using superposition principle of Quantum physics. The author in the experimental analysis of the article, for high-resolution imagery, compared three Quantum encoding techniques and image compression using a Quantum circuit and proposed a fourth Quantum encoding technique. For image size reduction, the author used a 200-image data set for analysis of the proposed fourth Quantum embedding method. By considering parameters such as performance, visual, and evaluation, the Author analyzes the tuning effect on compressed images on the Quantum circuit [14].

3.5.6 Quantum Fourier Transform

QFT is the Quantum version of the Discrete Fourier transform. When on the Quantum state amplitude probability shown in figure 11, DFT is applied by rearranging all the terms into a product form [11].

3.5.7 Differences and characteristics of available Quantum datasets

For the implementation of the algorithm Using IBM QISKIT Script and Matlab Code. The Quantum eigenvalue estimation circuit for the Hankel patterned Hamiltonian is implemented on an IBM Quantum simulator. The author also made an open-access QISKIT script for this simulation available. For easy reference for the understanding of gate-level implementation for a Hankel matrix using elementary Quantum gates, the author considered MATLAB codes

4 Discussion

A Quantum walk discussion is conducted in this paper to construct a Quantum artificial neural network to speed up the Quantum machine learning algorithm exponentially. All classical supervised and unsupervised machine learning algorithms are compared with Quantum-based supervised and unsupervised machine learning algorithms to determine the limits of existing machine learning techniques. It is observed that QPSO-TWSVM, a Quantum particle swarm optimization twin-wave support vector machine, helped to provide optimized classification accuracy because QPSO helps to find optimum parameters more precisely and has good global convergence that improves TWSVM classification accuracy and efficiency.

5 Conclusion and future directions

Quantum machine learning could be used to classify remote sensing images that could have been explored in this paper. Further research work on image classification using Quantum machine learning is necessary for research areas of remote sensing in which big data needs to be handled more conveniently. There is also lots of work also needs to be done to make Quantum machines more fault tolerant. Since some current Quantum machines support few qubits that is disadvantageous in many Quantum machine learning algorithm-based applications. Quantum neural network, a branch of Quantum machine learning can perform the classical counterparts which comes under one of the challenges of Quantum computing and has lots of research work that needs to be implemented.

Declarations

– Funding: The authors received no specific funding for this study.

– Conflict of interest: The authors declare that they have no conflicts of interest to report regarding the present study.
– Ethics approval: Approved
– Consent to participate: Not applicable
– Consent for publication: Not applicable
– Availability of data and materials: Not applicable
– Code availability: Not applicable
– Authors' contributions: All authors contributed equally

References

1. Yehui Tang, Junchi Yan, Guoqiang Hu, Baohua Zhang, and Jinzan Zhou. Recent progress and perspectives on quantum computing for finance. *Service Oriented Computing and Applications*, 16:227 – 229, 2022. URL https://api.semanticscholar.org/CorpusID:252572179.
2. Xiwei Yao, Hengyan Wang, Zeyang Liao, Ming-Cheng Chen, Jian-Wei Pan, Jun Yu Li, Kechao Zhang, Xingcheng Lin, Zhehui Wang, Zhihuan Luo, Wenqiang Zheng, Jianzhong Li, Mei sheng Zhao, Xinhua Peng, and Dieter Suter. Quantum image processing and its application to edge detection: Theory and experiment. *arXiv: Quantum Physics*, 2017. URL https://api.semanticscholar.org/CorpusID:119205332.
3. Hailing Liu, Chao-Hua Yu, Yusen Wu, Shijie Pan, Sujuan Qin, Fei Gao, and Qiaoyan Wen. Quantum algorithm for logistic regression. *arXiv: Quantum Physics*, 2019. URL https://api.semanticscholar.org/CorpusID:182952877.
4. Jisun Park, Yong Jin, Seoungjae Cho, Yunsick Sung, and Kyungeun Cho. Advanced machine learning for gesture learning and recognition based on intelligent big data of heterogeneous sensors. *Symmetry*, 11:929, 2019. URL https://api.semanticscholar.org/CorpusID:198942233.
5. Nagamani Tenali and Gatram Rama Mohan Babu. Hqdc-net: Hybrid quantum dilated convolution neural network for detecting covid-19 in the context of big data analytics. *Multimedia Tools and Applications*, pages 1 – 27, 2023. URL https://api.semanticscholar.org/CorpusID:258674970.
6. Yousef Ameen Esmail Ahmed, Biao Yue, Zongliang Gu, and Jingyu Yang. An overview: Big data analysis by deep learning and image processing. *International Journal of Quantum Information*, 2023. URL https://api.semanticscholar.org/CorpusID:257199626.
7. Xiang ming Gao and Ruijun Li. A kernel support vector machine for big data mining based on buffered k-d tree. 2016. URL https://api.semanticscholar.org/CorpusID:248121393.
8. Phillip A. Kerger and Ryoji Miyazaki. Quantum image denoising: a framework via boltzmann machines, qubo, and quantum annealing. *Frontiers in Computer Science*, 2023. URL https://api.semanticscholar.org/CorpusID:259847393.
9. Zexin Huang. Research on computer big data technology in the creation of virtual reality animation character dance movement system. *Proceedings of the 7th International Conference on Cyber Security and Information Engineering*, 2022. URL https://api.semanticscholar.org/CorpusID:253120747.
10. Lin Wang, Haiyan Zhang, and Guoliang Yuan. Big data and deep learning-based video classification model for sports. *Wireless Communications and Mobile Computing*, 2021. URL https://api.semanticscholar.org/CorpusID:241544519.
11. José David Martín-Guerrero and Lucas Lamata. Quantum machine learning. In *The European Symposium on Artificial Neural Networks*, 2020. URL https://api.semanticscholar.org/CorpusID:196181421.
12. Mengjun Huang. On the role of artificial intelligence and big data in visual arts. *Frontiers in Science and Engineering*, 2022. URL https://api.semanticscholar.org/CorpusID:255639784.
13. Wei Hu. Empirical analysis of a quantum classifier implemented on ibm's 5q quantum computer. 2018. URL https://api.semanticscholar.org/CorpusID:208869477.
14. Yahui Hu, Qiushi Xi, Lei Xiao, Qiqing Fang, and Yamin Hu. Research on image analysis and processing technology based on big data technology. *2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, pages 1383–1386, 2023. URL https://api.semanticscholar.org/CorpusID:258070206.
15. Gabriele Cavallaro, Morris Riedel, Thomas Lippert, and Kristel Michielsen. Hybrid quantum-classical workflows in modular supercomputing architectures with the julich unified infrastructure for quantum computing. *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 4149–4152, 2022. URL https://api.semanticscholar.org/CorpusID:252589081.
16. Naim M. Ajlouni, Adem Özyavas, Mustafa Takaoğlu, Faruk Takaoglu, and Firas Ajlouni. Medical image diagnosis based on adaptive hybrid quantum cnn. *BMC Medical Imaging*, 23, 2023. URL https://api.semanticscholar.org/CorpusID:261708198.
17. Steven E Dilsizian and Eliot L. Siegel. Artificial intelligence in medicine and cardiac imaging: Harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. *Current Cardiology Reports*, 16:1–8, 2013. URL https://api.semanticscholar.org/CorpusID:20958058.
18. J. P. Medlin Julia and Dennis Bennet. Efficient machine learning algorithms for medical big data analysis. *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, pages 795–800, 2022. URL https://api.semanticscholar.org/CorpusID:249803309.
19. Benno Torgler. Big data, artificial intelligence, and quantum computing in sports. 2020. URL https://api.semanticscholar.org/CorpusID:226518595.
20. Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd. Quantum support vector machine for big feature and big data classification. *Physical review letters*, 113 13:130503, 2013. URL https://api.semanticscholar.org/CorpusID:5503025.
21. Steve Blair and Jon Cotter. An analysis of data processing for big data analytics. *Journal of Computing and Natural Science*, 2021. URL https://api.semanticscholar.org/CorpusID:245732674.
22. Nan Jiang. Role of digital image processing in image art under the background of big data. *IOP Conference Series: Materials Science and Engineering*, 750, 2020. URL https://api.semanticscholar.org/CorpusID:216336736.
23. Nana Liu and Patrick Rebentrost. Quantum machine learning for quantum anomaly detection. *Physical Review A*, 97, 2017. URL https://api.semanticscholar.org/CorpusID:55055029.
24. Jin Wang, Yaqiong Yang, Tian Wang, Robert Simon Sherratt, and Jingyu Zhang. Big data service architecture: A survey. *Journal of Internet Technology*, 21:

393–405, 2020. URL https://api.semanticscholar.org/CorpusID:216326835.
25. Lov K. Grover. Quantum mechanics helps in searching for a needle in a haystack. *Physical Review Letters*, 79: 325–328, 1997. doi: 10.1103/PhysRevLett.79.325.
26. Phillip Kerger and Ryoji Miyazaki. Quantum image denoising: A framework via boltzmann machines, qubo, and quantum annealing. *arXiv preprint arXiv:2307.06542*, 2023.
27. Mihai Datcu. Machine learning meets computational imaging: Big data analytics for earth observation. 2018. URL https://api.semanticscholar.org/CorpusID:69585470.
28. Dawei Sun, Guangyan Zhang, Weimin Zheng, and Keqin Li. Key technologies for big data stream computing. In *Big Data*, 2015. URL https://api.semanticscholar.org/CorpusID:9055712.
29. Duan Jingbo. Big data classification model and algorithm based on double quantum particle swarm optimization. *2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT)*, pages 1309–1313, 2023. URL https://api.semanticscholar.org/CorpusID:259100936.
30. Young-Sik Jeong and Jong Hyuk Park. Advanced big data analysis, artificial intelligence & communication systems. *J. Inf. Process. Syst.*, 15:1–6, 2019. URL https://api.semanticscholar.org/CorpusID:83459446.
31. Gagan Sharma and Sunil Patil. Big data analysis for revenue and sales prediction using support vector regression with auto-regressive integrated moving average. *SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology*, 2023. URL https://api.semanticscholar.org/CorpusID:261166626.
32. AS Kussainov, AT Karimova, SG Kussainov, and NY Pya. Immediate challenges faced by the quantum computing in time series analysis. . *()*, 44(1):101–105, 2013.
33. Chen Ding, Tianyi Bao, and Heliang Huang. Quantum-inspired support vector machine. *IEEE Transactions on Neural Networks and Learning Systems*, 33:7210–7222, 2019. URL https://api.semanticscholar.org/CorpusID:195316662.
34. Yufei Ding and Ali Javadi-Abhari. Quantum and post-moore's law computing. *IEEE Internet Computing*, 26:5–6, 2022. URL https://api.semanticscholar.org/CorpusID:246657331.
35. Li Jian Xu, Xiaoyu Zhang, Jing Wang, Ming Li, Ling Jian, and Shuqian Shen. Variational quantum support vector machine based on hadamard test. *Communications in Theoretical Physics*, 74, 2022. URL https://api.semanticscholar.org/CorpusID:247928366.
36. Tang Ting. Research on intelligent image scrambling transform encryption algorithm based on big data analysis. *2021 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, pages 720–723, 2021. URL https://api.semanticscholar.org/CorpusID:237403177.
37. Md. Ershadul Haque, Manoranjan Paul, Anwaar Ulhaq, and Tanmoy Debnath. Advanced quantum image representation and compression using a dct-efrqi approach. *Scientific Reports*, 13, 2023. URL https://api.semanticscholar.org/CorpusID:257485968.
38. Nico Vervliet, Otto Debals, Laurent Sorber, and Lieven De Lathauwer. Breaking the curse of dimensionality using decompositions of incomplete tensors: Tensor-based scientific computing in big data analysis. *IEEE Signal Processing Magazine*, 31:71–79, 2014. URL https://api.semanticscholar.org/CorpusID:8051815.
39. Juan Zhong and Bo He. Application of big data analysis and image processing technology in athletes training based on intelligent machine vision technology. *Lecture Notes on Data Engineering and Communications Technologies*, 2021. URL https://api.semanticscholar.org/CorpusID:244206123.
40. Vladimir Hahanov, Eugenia Litvinova, Svetlana Chumachenko, Tetiana Soklakova, and Irina V. Hahanova. Big data quantum computing. 2018. URL https://api.semanticscholar.org/CorpusID:65342779.
41. Fan Fan, Yilei Shi, Tobias Guggemos, and Xiaoxiang Zhu. Hybrid quantum-classical convolutional neural network model for image classification. *IEEE transactions on neural networks and learning systems*, PP, 2023. URL https://api.semanticscholar.org/CorpusID:262054805.
42. Jianqing Fan, Fang Han, and Han Liu. Challenges of big data analysis. *National science review*, 1 2:293–314, 2013. URL https://api.semanticscholar.org/CorpusID:12483505.
43. Manju Kumari and Shailender Gupta. Performance comparison between chaos and quantum-chaos based image encryption techniques. *Multimedia Tools and Applications*, 80:33213 – 33255, 2021. URL https://api.semanticscholar.org/CorpusID:237090261.
44. Wafaa S. Albaldawi, Rafah M. Almuttairi, and Mehdi Ebady Manaa. Big data analysis for healthcare application using minhash and machine learning in apache spark framework. *2022 International Conference on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pages 1–7, 2022. URL https://api.semanticscholar.org/CorpusID:250119376.
45. Yasuhiro Ohkura, Takahiko Satoh, and Rodney Van Meter. Simultaneous execution of quantum circuits on current and near-future nisq systems. *IEEE Transactions on Quantum Engineering*, 3:1–10, 2022. doi: 10.1109/TQE.2022.3164716.

Operations Management Research
https://doi.org/10.1007/s12063-024-00462-8

RETRACTION NOTE

Retraction Note: Modelling product lines diffusion: a framework incorporating competitive brands for sustainable innovations

Deepti Aggrawal[1] · Adarsh Anand[2] · Gunjan Bansal[2] · Gareth H. Davies[3] · Parisa Maroufkhani[4,5] · Yogesh K. Dwivedi[6,7]

© The Author(s) 2024

Retraction Note: Operations Management Research (2022) 15:760–772
https://doi.org/10.1007/s12063-022-00260-0

The Publisher has retracted this article in agreement with the Editor-in-Chief. The article was submitted to be part of a guest-edited issue. An investigation by the publisher found a number of articles, including this one, with a number of concerns, including but not limited to compromised editorial handling and peer review process, inappropriate or irrelevant references or not being in scope of the journal or guest-edited issue. Based on the investigation's findings the publisher, in consultation with the Editor-in-Chief therefore no longer has confidence in the results and conclusions of this article.

Author Yogesh K. Dwivedi disagrees with this retraction. The other authors have not responded to correspondence regarding this retraction.

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/.

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

The online version of the original article can be found at https://doi.org/10.1007/s12063-022-00260-0.

✉ Deepti Aggrawal
deeptiaggrawal@dtu.ac.in

✉ Adarsh Anand
adarsh.anand86@gmail.com

✉ Gunjan Bansal
gunjan.1512@gmail.com

✉ Gareth H. Davies
g.h.davies@swansea.ac.uk

✉ Parisa Maroufkhani
parisa.maroufkhani@gmail.com

✉ Yogesh K. Dwivedi
y.k.dwivedi@swansea.ac.uk

[1] University School of Management and Entrepreneurship, Delhi Technological University, Delhi, India

[2] Department of Operational Research, University of Delhi, Delhi, India

[3] School of Management, Swansea University, Swansea, UK

[4] The University of Waikato Joint Institute, Zhejiang University City College, Hangzhou, PR China

[5] Zhejiang University City College, Hangzhou, People's Republic of China

[6] Emerging Markets Research Centre (EMaRC), School of Management, Swansea University, Room #323, Bay Campus, Fabian Bay, Swansea, Wales SA1 8EN, UK

[7] Department of Management, Symbiosis Institute of Business Management, Pune & Symbiosis International (Deemed University), Pune, Maharashtra, India

Published online: 13 February 2024

Physics of Plasmas

RESEARCH ARTICLE | FEBRUARY 02 2024

Role of nonlinear structures and associated turbulence generation dayside magnetosphere reconnection sites

Neha Pathak ; Suresh C. Sharma ; Garima Patel ; R. P. Sharma

Check for updates

Phys. Plasmas 31, 022104 (2024)
https://doi.org/10.1063/5.0172154

View Online
Export Citation
CrossMark

21 February 2024 06:12:46



AIP Publishing

Role of nonlinear structures and associated turbulence generation dayside magnetosphere reconnection sites

Cite as: Phys. Plasmas **31**, 022104 (2024); doi: 10.1063/5.0172154
Submitted: 13 August 2023 · Accepted: 12 January 2024 ·
Published Online: 2 February 2024

View Online Export Citation CrossMark

Neha Pathak,[1,a)] Suresh C. Sharma,[1] Garima Patel,[2] and R. P. Sharma[2]

AFFILIATIONS

[1]Department of Applied Physics, Delhi Technological University, Shahbad Daulatpur, Bawana Road, New Delhi 110042, India
[2]Department of Energy Science and Engineering, Indian Institute of Technology Delhi, Delhi 110016, India

[a)]Author to whom correspondence should be addressed: npsneha90@gmail.com

ABSTRACT

A numerical simulation is implemented to investigate the role of whistler wave in the presence of the ponderomotive force-driven density modification and the magnetic field perturbation due to preexisting magnetic islands. The dynamical equation governing the whistler waves is derived and solved using numerical methods to assess their role. The simulation produced results that revealed the nonlinear structures, confirming the system's turbulent characteristics. Moreover, we utilized a semi-analytical model, applying the paraxial approximation, to estimate the scale size of the current sheet, which was found to be between $2\lambda_e$ and λ_e. This measurement provides evidence of the presence of a superthin electron current sheet within reconnection sites. Qualitative consistency is established by comparing the simulation results with findings reported in the literature.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0172154

21 February 2024 06:12:46

I. INTRODUCTION

Plasma turbulence is an omnipresent phenomenon observed in various astrophysical contexts and plays a central role in several areas, including the solar corona,[1] the solar wind,[2] and active regions within Earth's magnetosphere.[3–8] It assumes a pivotal role in enabling the transfer and dissipation of energy across different scales in space, astrophysics, and laboratory plasmas. While magnetic field annihilation, driven by magnetic reconnection mechanisms as detailed in studies,[9,10] serves as a significant contributor to turbulence generation.

Furthermore, mounting evidence from several studies suggests that turbulence exhibits a close interconnection with the magnetic reconnection, as exemplified in research.[4,11–13] Turbulence in preexisting magnetic reconnection is often described as a magnetic reconnection-driven turbulence.[14] Delving deeper into the nature of turbulence, it is valuable to examine its specific components. Past observations have revealed that magnetic reconnection triggers the generation of a diverse spectrum of waves across a wide frequency range. These waves encompass lower hybrid waves, Langmuir waves, electrostatic solitary waves, and whistler waves, as demonstrated in studies.[15–18] The nonlinear interaction between waves facilitates the transfer of energy across different scales and influences the exchange of energy between electromagnetic fields and particles. However, a comprehensive understanding of their coupling with other wave types, as well as the nature of their nonlinear interaction, remains an open question.

For instance, within the magnetosphere, ultra-low-frequency waves have been observed to modulate whistler and electron cyclotron harmonic (ECH) waves.[19] Additionally, kinetic Alfvén waves have been noted to modulate whistler waves.[20] In specific scenarios, oblique-propagating whistler waves have been found to generate parallel beams that, in turn, drive Langmuir or electron acoustic waves.[21] Several theories have been put forth in the literature to elucidate the process of particle acceleration in turbulence driven by magnetic reconnection. These theories encompass the influence of several factors, including the parallel electric field, the Fermi acceleration, and the betatron mechanisms. Additionally, the interaction between magnetic islands emerges as a pivotal mechanism in this context.[22]

Whistler waves have emerged as significant contributors in the realm of turbulence within space plasmas. Previous studies have recognized their crucial role in mediating energy transfer across various scales. The interaction with other plasma waves (including ion acoustic waves, kinetic Alfvén waves, and electron cyclotron waves, to name a few) can lead to a rich array of nonlinear phenomena, such as wave–wave coupling, wave–particle interactions, and the generation of

Published under an exclusive license by AIP Publishing

secondary waves.[23–25] However, a systematic understanding if waves are present in preexisting reconnection sites how they affect the turbulence generation is still lacking.

In this study, we employ a wave based numerical model to investigate turbulence within the dayside magnetosphere reconnection sites. Additionally, a wave based semi-analytical model is employed to elucidate the physics of the evolution of a superthin electron current sheet at electron scales. The present numerical model deals with the nonlinear evolution of whistler waves through the preexisting reconnection sites. For this study, the dynamical evolution equations are derived by taking into account the ponderomotive force-driven density modification and the magnetic field perturbation due to magnetic island. The present model has nonlinear effects associated with whistler waves, with the aim of advancing our understanding of these intriguing phenomena. Understanding these nonlinear interactions is essential for comprehending the overall behavior of space plasmas, particularly in regions like the magnetosphere and the solar wind, where multiple wave modes coexist. These interactions have implications for energy transfer processes, particle acceleration, and the overall evolution of plasma turbulence.

In Sec. II, we describe the derivation algebra steps for 3D whistler waves. In Sec. III, we introduce the simulation method employed in this study. Moving on to Sec. IV, we discuss the semi-analytical model applied in the present work. In Sec. V, we present the simulation results, and finally, in Sec. VI, we provide a summary and conclusion for this study.

II. BASIC EQUATIONS

Whistler wave is an electromagnetic wave in the frequency range of $\omega_{ci} \ll \omega \ll \omega_{ce}$, where ω_{ci} is the ion gyrofrequencies and ω_{ce} is the electron gyrofrequencies. This wave's behavior can be effectively described with spatial and temporal scales smaller than that of ions. It is reasonable to assume in this analysis that ions are motionless.[26] Researchers have widely analyzed the linear and nonlinear characteristics of whistler waves under these specific conditions.[27–31] In this study, we investigate the dispersion relation of whistler wave, taking into account the propagation vector $\boldsymbol{k} = k_x\hat{x} + k_y\hat{y} + k_z\hat{z}$ and the magnetic field $\boldsymbol{B} = B_{0z}\hat{z} + B_{0y}\hat{y}$, where B_{0z} represents the background magnetic field in the z direction and B_{0y} denotes an additional transverse magnetic field in the y direction due to the presence of a preexisting magnetic island. In the context of whistler waves, electron fluid currents are of significance, since here the ions are assumed as the motionless neutralizing background.[26,32] Consequently, we will focus on equations relevant to electron dynamics, including those governing electron fluid momentum, currents, and Faraday's law.

If we assume a collisionless plasma with no thermal motions, the equation of motion for the electron can be written as

$$\frac{\partial \boldsymbol{v}_e}{\partial t} = -\frac{e}{\mathrm{m}_e}\boldsymbol{E} - \frac{e}{m_e c}(\boldsymbol{v}_e \times \boldsymbol{B}), \tag{1}$$

$$\text{Ampere's law, } \boldsymbol{\nabla} \times \boldsymbol{B} = \frac{4\pi}{c}\boldsymbol{j}, \tag{2}$$

$$\text{Faraday's law, } \boldsymbol{\nabla} \times \boldsymbol{E} = -\frac{1}{c}\frac{\partial \boldsymbol{B}}{\partial t}. \tag{3}$$

Here, m_e is the electron mass density, e is the electron charge, $\boldsymbol{v}_e$ is the electron fluid velocity, and c is the speed of light in a vacuum. $\boldsymbol{E}$ and $\boldsymbol{B}$ represent the electric and the magnetic fields. $\boldsymbol{j}$ is the electron current density. Due to the significant difference in magnitude between the electron plasma frequency (ω_{pe}) and the electron cyclotron frequency (ω_{ce}) in Earth's magnetosphere, the displacement current is typically ignored in Ampère's law when analyzing whistler wave properties.[32] It has been reported[33] that a significant portion of the fluctuation energy in turbulent plasmas is carried by the magnetic field $(\boldsymbol{B})$. However, since $\boldsymbol{B}$ cannot perform work on the plasma, the exchange of energy between electromagnetic fields and particles occurs primarily through the electric field $(\boldsymbol{E})$. In collisionless plasmas, $\boldsymbol{E}$ is governed by a generalized Ohm's law, derived from the momentum equations of electrons and/or ions

$$\boldsymbol{E} = \frac{1}{nec}(\boldsymbol{j} \times \boldsymbol{B}) - \frac{1}{c}(\boldsymbol{v}_e \times \boldsymbol{B}) + \frac{m_e}{ne^2}\frac{\partial \boldsymbol{j}}{\partial t}. \tag{4}$$

On the right side, the first term corresponds to the Hall term arising due to the differential motion between electrons and ions. The second term is associated with electrons being fully frozen into the magnetic field. Since we are considering that magnetic field lines are fully frozen into the electron and ions are motionless i.e., $\boldsymbol{j} = -en\boldsymbol{v}_e$ and here, n is the density. In this analysis, we have focused on a collisionless plasma with no thermal motions. As a result, we have neglected the electron thermal pressure and the viscosity terms in Eq. (4).

To derive the dispersion relation of the whistler wave, we begin by taking the curl of the generalized Ohm's law, as depicted in Eq. (4), and subsequently substituting the term related to the induced electric field with Faraday's law. The electron current density can be associated with the rotational magnetic field through Ampère's law, $\boldsymbol{j} = \frac{c}{4\pi}\boldsymbol{\nabla} \times \boldsymbol{B}$. Taking this into account, we can rewrite the magnetic induction equation as follows:

$$\frac{\partial \boldsymbol{B}}{\partial t} = -\frac{m_e c^2}{4\pi ne^2}\boldsymbol{\nabla} \times \frac{\partial(\boldsymbol{\nabla} \times \boldsymbol{B})}{\partial t} - \frac{c}{4\pi ne}\boldsymbol{\nabla} \times [(\boldsymbol{\nabla} \times \boldsymbol{B}) \times \boldsymbol{B}] + \boldsymbol{\nabla} \times (\boldsymbol{v}_e \times \boldsymbol{B}). \tag{5}$$

Additionally, we apply the vector identities $\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \boldsymbol{A}) = \boldsymbol{\nabla}(\boldsymbol{\nabla}.\boldsymbol{A}) - \boldsymbol{\nabla}^2\mathbf{A}$ and $\boldsymbol{\nabla} \times (\boldsymbol{A} \times \boldsymbol{B}) = \boldsymbol{A}(\boldsymbol{\nabla}.\boldsymbol{B}) - \boldsymbol{B}(\boldsymbol{\nabla}.\boldsymbol{A}) + (\boldsymbol{B}.\boldsymbol{\nabla})\boldsymbol{A} - (\boldsymbol{A}.\boldsymbol{\nabla})\boldsymbol{B}$ to the right side of Eq. (5). It is important to note that we consider $\boldsymbol{\nabla}.\boldsymbol{B} = 0$ and $\boldsymbol{\nabla}.\boldsymbol{v}_e = 0$ (for an incompressible fluid).

The right side terms in Eq. (5) can be expressed in the following forms: the first term is $\frac{m_e c^2}{4\pi ne^2}\boldsymbol{\nabla} \times \frac{\partial(\boldsymbol{\nabla} \times \boldsymbol{B})}{\partial t} = -\lambda_e^2\frac{\partial(\boldsymbol{\nabla}^2\boldsymbol{B})}{\partial t}$, the second term is $\frac{\sqrt{m_i}c}{\sqrt{4\pi nm_i}\sqrt{4\pi ne}}\boldsymbol{\nabla} \times [(\boldsymbol{\nabla} \times \boldsymbol{B}) \times \boldsymbol{B}] = \frac{\lambda_i}{\sqrt{4\pi nm_i}}(\boldsymbol{B}.\boldsymbol{\nabla})(\boldsymbol{\nabla} \times \boldsymbol{B})$, and the third term particularly in the limit of $k\lambda_i > 1$ simplifies as[34] $\boldsymbol{\nabla} \times (\boldsymbol{v}_e \times \boldsymbol{B}) = (\boldsymbol{B}.\boldsymbol{\nabla})\boldsymbol{v}_e$. The magnitude of the second and the third terms differs significantly, with orders of magnitude separating them

$$\boldsymbol{\nabla} \times (\boldsymbol{v}_e \times \boldsymbol{B}) = (\boldsymbol{B}.\boldsymbol{\nabla})\boldsymbol{v}_e \sim b^2/l,$$
$$\frac{\lambda_i}{\sqrt{4\pi nm_i}}\boldsymbol{\nabla} \times [(\boldsymbol{\nabla} \times \boldsymbol{B}) \times \boldsymbol{B}] = \frac{\lambda_i}{\sqrt{4\pi nm_i}}(\boldsymbol{B}.\boldsymbol{\nabla})(\boldsymbol{\nabla} \times \boldsymbol{B}) \sim \lambda_i b^2/l^2\sqrt{4\pi nm_i}.$$

Here, l is the scale of interest and we assume $v \sim b$. Therefore, Eq. (5) reduces to the following equation:

$$\frac{\partial \boldsymbol{B}}{\partial t} = \lambda_e^2\frac{\partial(\boldsymbol{\nabla}^2\boldsymbol{B})}{\partial t} - \frac{\lambda_i}{\sqrt{4\pi nm_i}}(\boldsymbol{B}.\boldsymbol{\nabla})(\boldsymbol{\nabla} \times \boldsymbol{B}). \tag{6}$$

The linearization of Eq. (6) about a magnetic field $\boldsymbol{B} = B_{0z}\hat{z} + B_{0y}\hat{y} + B_1(x, y, z, t)$ yields the following equation:

21 February 2024 06:12:46

Published under an exclusive license by AIP Publishing

$$-\omega B_1 - \omega\lambda_e^2 k^2 B_1 = -\frac{\lambda_i}{\sqrt{4\pi n m_i}}\left(k_z B_{0z} + k_y B_{0y}\right) k B_1. \quad (7)$$

Here, ω is the frequency of the whistler wave and B_1 is associated with the whistler wave magnetic fields. Equation (8) represents the dispersion relation of the whistler wave having a propagation in a medium background magnetic field (B_{0z}), which is along the z direction and also there is the magnetic field B_{0y} due to the presence of a preexisting magnetic island

$$\omega = \frac{k\lambda_i\left(k_z B_{0z} + k_y B_{0y}\right)}{\sqrt{4\pi n m_i}\left(1 + k^2\lambda_e^2\right)}. \quad (8)$$

If the effect of the preexisting magnetic island is not present, one gets the usual dispersion relation of the whistler wave under the condition[35] $k\lambda_i > 1$

$$\omega\left(1 + k^2\lambda_e^2\right) = \lambda_i v_A k k_z. \quad (9)$$

Here, $\lambda_i = \sqrt{m_i}c/\sqrt{4\pi ne}$ is the ion inertial length, $\lambda_e = \sqrt{m_e}c/\sqrt{4\pi ne}$ is the electron inertial length, and $v_A = B_{0z}/\sqrt{4\pi n m_i}$ is Alfvén speed.

The whistler wave magnetic field can be represented as $B_1 = \tilde{B}_1 \exp[i(\boldsymbol{k}.\boldsymbol{r} - \omega t)]$. If the wave propagates in three-dimensions $\boldsymbol{k}.\boldsymbol{r}$ becomes, $\boldsymbol{k}.\boldsymbol{r} = k_x x + k_y y + k_z z$. We get the governing equation in terms of the vector potential ($\tilde{A}_z$) by applying the inverse Fourier transform of Eq. (8) and coupling with the relation $\vec{B}_{1\perp} = \hat{z} \times \boldsymbol{\nabla}\tilde{A}_z$ as given by[36]

$$\frac{\partial^2\tilde{A}_z}{\partial t^2} + \lambda_e^4\frac{\partial^6\tilde{A}_z}{\partial t^2\partial x^4} + \lambda_e^4\frac{\partial^6\tilde{A}_z}{\partial t^2\partial y^4} + \lambda_e^4\frac{\partial^6\tilde{A}_z}{\partial t^2\partial z^4} + 2\lambda_e^4\boldsymbol{\nabla}_\perp^2\frac{\partial^4\tilde{A}_z}{\partial t^2\partial z^2}$$
$$- 2\lambda_e^2\frac{\partial^2}{\partial t^2}\left(\boldsymbol{\nabla}_\perp^2 + \frac{\partial^2}{\partial z^2}\right)\tilde{A}_z + \lambda_i^2 v_A^2\boldsymbol{\nabla}_\perp^2\frac{\partial^2\tilde{A}_z}{\partial z^2} + \lambda_i^2 v_A^2\frac{\partial^4\tilde{A}_z}{\partial z^4}$$
$$+ \lambda_i^2 v_A^2\left(\delta\tilde{A}_z\right)^2\boldsymbol{\nabla}_\perp^2\frac{\partial^2\tilde{A}_z}{\partial y^2} + \lambda_i^2 v_A^2\left(\delta\tilde{A}_z\right)^2\frac{\partial^4\tilde{A}_z}{\partial y^2\partial z^2}$$
$$+ 2\lambda_i^2 v_A^2\left(\delta\tilde{A}_z\right)\boldsymbol{\nabla}_\perp^2\frac{\partial^2\tilde{A}_z}{\partial y\partial z} + 2\lambda_i^2 v_A^2\left(\delta\tilde{A}_z\right)\frac{\partial^4\tilde{A}_z}{\partial y\partial z^3} = 0. \quad (10)$$

In Eq. (10), $\tilde{A}_z$ represents the whistler wave vector potential, $\delta\tilde{A}_z$ is the perturbation arising due to the presence of magnetic islands. Here, we have used the relation of the vector potential and the magnetic field $\vec{B}_{1\perp} = \hat{z} \times \boldsymbol{\nabla}\tilde{A}_z$, i.e., $\frac{B_{0y}}{B_{0z}} = \frac{\delta\tilde{A}_z}{A_0}$. The present investigation focuses on the nonlinear progression of whistler waves within a modified preexisting magnetic island influenced by the whistler wave field. In this study, we derive the dynamic equation, considering the effects of ponderomotive force-induced density changes and perturbations in the magnetic field resulting from the presence of the magnetic island. Assuming the envelope solution $\tilde{A}_z = A_z(x, y, z)\exp[-i(\omega t - k_x x - k_y y - k_z z)]$ (here, $A_z(x, y, z)$ is the slowly varying amplitude) and substituting this solution in Eq. (10). We get the following equation:

$$\frac{-2i\omega\left(1 + 9k^4\lambda_e^4 + 18k_\perp^2 k_z^2\lambda_e^4 + 6k^2\lambda_e^2\right)}{\lambda_i^2 v_A^2 k^2 k_z^2}\frac{\partial A_z}{\partial t} + \frac{2ik_\perp\left(2k^2\omega^2\lambda_e^4 + 2\omega^2\lambda_e^2 - \lambda_i^2 v_A^2 k_z^2\right)}{\lambda_i^2 v_A^2 k^2 k_z^2}\left(\frac{\partial A_z}{\partial x} + \frac{\partial A_z}{\partial y}\right)$$
$$- \frac{2ik_z\left(-2k^2\omega^2\lambda_e^4 - 2\omega^2\lambda_e^2 + k^2\lambda_i^2 v_A^2 + \lambda_i^2 v_A^2 k_z^2\right)}{\lambda_i^2 v_A^2 k^2 k_z^2}\frac{\partial A_z}{\partial z} + \frac{4\omega^2 k_\perp^2\lambda_e^4}{\lambda_i^2 v_A^2 k^2 k_z^2}\left(\frac{\partial^2 A_z}{\partial x^2} + \frac{\partial^2 A_z}{\partial y^2}\right) - \frac{\left(4\lambda_i^2 v_A^2 - 4\omega^2\lambda_e^4\right)}{\lambda_i^2 v_A^2 k^2}\frac{\partial^2 A_z}{\partial z^2}$$
$$+ \frac{\left(8k_\perp k_z\omega^2\lambda_e^4 - 4k_\perp k_z\lambda_i^2 v_A^2\right)}{\lambda_i^2 v_A^2 k^2 k_z^2}\left(\frac{\partial^2 A_z}{\partial x\partial z} + \frac{\partial^2 A_z}{\partial y\partial z}\right) - \frac{k_y^2}{k_z^2}\left(\delta\tilde{A}_z\right)^2 A_z - 2\frac{k_y}{k_z}\left(\delta\tilde{A}_z\right)A_z + \frac{\delta n}{n}A_z = 0. \quad (11)$$

Here, the field perturbation caused by the preexisting magnetic island can be written as[37] $\delta\tilde{A}_z = \left(-\frac{x^2}{2} + A_z\cos\left(k_y y\right)\right)$. We examine the scenario where whistler waves propagate in the presence of preexisting magnetic islands, alongside background density perturbations resembling ion acoustic fluctuations. Specifically, we are considering the perturbations in the magnetic field $\left(\delta\tilde{A}_z\right)$, caused by the preexisting magnetic islands and perturbations in the number density $(\delta n/n)$, due to the presence of ion acoustic waves within the background plasma. In this context, we are assuming that the ion acoustic wave propagates in the z-direction. To derive the dynamical equation for the ion acoustic wave, we utilize the equation of motion for both electrons and ions, along with the continuity equation, as follows:

$$\left(\frac{\partial^2}{\partial t^2} - c_s^2\frac{\partial^2}{\partial z^2}\right)\frac{\delta n}{n_0} = -\frac{\partial}{\partial z}\left(\frac{F_{ez} + F_{iz}}{m_i}\right). \quad (12)$$

Here, c_s is the ion sound speed, and F_{ez} and F_{iz} are the z-components of the ponderomotive forces due to whistler wave.

We can drive expressions for F_{ez} and F_{iz} by employing the equation of motion and Faraday's law

$$F_{ez} = -\frac{e^2}{2m_e c^2}\frac{\partial|A_z|^2}{\partial z},\ F_{iz} = -\frac{e^2}{2m_i c^2}\frac{\partial|A_z|^2}{\partial z}$$

When considering the adiabatic response of density,[36,38–40] the dynamical equation for the ion acoustic wave can be written as follows:

$$\frac{\partial^2(\delta n/n_0)}{\partial z^2} = \frac{e^2}{2m_e c^2 c_s^2}\left(\frac{1}{m_e} + \frac{1}{m_i}\right)\frac{\partial^2|A_z|^2}{\partial z^2}. \quad (13)$$

Equation (13) shows the relationship between density perturbation and vector potential, i.e., $\frac{\delta n}{n} = |A_z|^2$, here $A_N = \left(\frac{2m_e^2 m_i c^2 c_s^2}{e^2(m_e + m_i)}\right)^{1/2}$ is the normalization parameter of the vector potential A_z. To explore the combined impact of both nonlinear effects in a whistler wave dynamical equation, we have included both perturbation effects into Eq. (11)

21 February 2024 06:12:46

Published under an exclusive license by AIP Publishing

$$-i\frac{\partial A_z}{\partial t}+iw_1\left(\frac{\partial A_z}{\partial x}+\frac{\partial A_z}{\partial y}\right)+i\frac{\partial A_z}{\partial z}+\frac{\partial^2 A_z}{\partial x^2}+\frac{\partial^2 A_z}{\partial y^2}$$
$$-w_2\frac{\partial^2 A_z}{\partial z^2}+w_3\left(\frac{\partial^2 A_z}{\partial x\partial z}+\frac{\partial^2 A_z}{\partial y\partial z}\right)-w_4\left(-\frac{x^2}{2}+A_z\cos(k_y y)\right)^2 A_z$$
$$-w_5\left(-\frac{x^2}{2}+A_z\cos(k_y y)\right)A_z+|A_z|^2A_z=0. \quad (14)$$

Here, $w_1=\frac{2k_\perp\left(2k^2\omega^2\lambda_e^4+2\omega^2\lambda_e^2-\lambda_i^2v_A^2k_z^2\right)}{\lambda_i^2v_A^2k^2k_z^2x_n}$, $w_2=\frac{\left(4\lambda_i^2v_A^2-4\omega^2\lambda_e^4\right)}{\lambda_i^2v_A^2k^2z_n^2}$, $w_3=\frac{\left(8k_\perp k_z\omega^2\lambda_e^4-4k_\perp k_z\lambda_i^2v_A^2\right)}{\lambda_i^2v_A^2k^2k_z^2x_nz_n}$, $w_4=\frac{k_y^2}{k_z^2}$, and $w_5=2\frac{k_y}{k_z}$ are constant used in the above Eq. (14).

The following normalization parameters are adopted to write equation (11) in normalized form, $t_n=\frac{2\omega\left(1+9k^4\lambda_e^4+18k_\perp^2k_z^2\lambda_e^4+6k^2\lambda_e^2\right)}{\lambda_i^2v_A^2k^2k_z^2}$, $x_n=y_n=\left(\frac{4\omega^2k_\perp^2\lambda_e^4}{\lambda_i^2v_A^2k^2k_z^2}\right)^2$, $z_n=\frac{2\left(2k^2\omega^2\lambda_e^4+2\omega^2\lambda_e^2-k^2\lambda_i^2v_A^2-\lambda_i^2v_A^2k_z^2\right)}{\lambda_i^2v_A^2k^2k_z}$.

Equation (14) describes the nonlinear behavior of a whistler wave propagating within a preexisting magnetic island and perturbed density.

III. NUMERICAL SIMULATION METHOD

The simulation is conducted using the pseudospectral method for space integration and finite difference method along with the predictor–corrector scheme for temporal integration.[41,42] For the temporal evolution, a finite difference method is employed with a time step $\Delta t=5\times10^{-4}$ (here, Δt is normalized with t_n), along with a predictor–corrector method. The simulation is conducted in a periodic box with dimensions $L_x=L_y=L_z=10\pi$, utilizing a grid with 128^3 points. In the current study, our algorithm development process involves two primary steps. In the first step, we have verified the accuracy of the code by adapting the algorithm to the nonlinear Schrödinger (NLS) equation and monitoring the invariants of this equation. Plasmon numbers, $N=\sum_k|A_K|^2$, are noted, which remain nearly constant with an accuracy extending up to five decimal points. Following this accuracy check, we subsequently applied the code to the model equations.

In the first step, the algorithm is applied to the nonlinear Schrödinger equation, where it computes various invariant parameters associated with this equation. Subsequently, the algorithm is adapted for use with dynamical equations and applied to investigate the observed Earth's magnetopause plasma parameters, as reported in the work of Chaston *et al.*[43] We have the flexibility to begin the simulation with various initial conditions, such as periodic, Gaussian, random, and X–O. Gaur and Sharma[44] have performed a simulation using different initial conditions. However, for our specific focus on studying turbulence generation in reconnection sites, we have chosen the magnetic field topology characterized by sequential X-points and O-points. This configuration is well-documented in the existing literature.[45,46] Biskamp and Welter[46] performed simulation and studied the dynamics of decaying the two-dimensional magnetohydrodynamic turbulence. For the present study, the initial profile of the field taken for the simulation at t = 0 is as follows:

$$A_z(x,y)=\cos(2x+2.3)+\cos(y+4.1). \quad (15)$$

The plasma parameters[43] used in the present work are $B_0=31\times10^{-5}$G, $n_0=21\text{cm}^{-3}$, $T_e=4.2\times10^5$K, $T_i=2.7\times10^6$K, $\beta=1.1$, $\lambda_e=1.1\times10^5$ cm, $\lambda_i=4.1\times10^6$ cm, $\rho_s=4.4\times10^6$ cm, $\omega_{ci}=2.9\,\text{s}^{-1}$, $\omega_{ce}=5\times10^3\text{s}^{-1}$, and $v_A=1.4\times10^7$cm/s.

IV. SEMI-ANALYTICAL MODEL

In the present manuscript, we have studied a steady-state semi-analytical model[47] within the paraxial regime ($r\ll r_0f$) to estimate the critical size of the localized structures. Here, r_0 denotes the transverse size of the whistler wave, and f represents the dimensionless beam width parameter. We have provided the derivation of the dynamical equation (14) in Sec. *II A*, which elucidates the nonlinear behavior of a whistler wave propagating within a preexisting magnetic island and perturbed density. Equation (11) in steady state, and by applying following the conditions $\partial_zA_z\ll k_zA_z$ and $\partial_rA_z\gg k_\perp A_z$ (paraxial approximations), can be written as

$$\frac{2ik_z\left(2k^2\omega^2\lambda_e^4+2\omega^2\lambda_e^2-k^2\lambda_i^2v_A^2-\lambda_i^2v_A^2k_z^2\right)}{\lambda_i^2v_A^2k^2k_z^2}\frac{\partial A_z}{\partial z}$$
$$+\frac{4\omega^2k_\perp^2\lambda_e^4}{\lambda_i^2v_A^2k^2k_z^2}\left(\frac{\partial^2A_z}{\partial x^2}+\frac{\partial^2A_z}{\partial y^2}\right)-\frac{k_y^2}{k_z^2}\left(\delta\tilde{A}_z\right)^2A_z$$
$$-2\frac{k_y}{k_z}\left(\delta\tilde{A}_z\right)A_z+\frac{\delta n}{n}A_z=0. \quad (16)$$

In a simplified form, Eq. (16) can be written as

$$\frac{2i}{k_z}\eta_1\frac{\partial A_z}{\partial z}+\lambda_e^2\eta_2\left(\frac{\partial^2A_z}{\partial x^2}+\frac{\partial^2A_z}{\partial y^2}\right)-\eta_3\left(-\frac{x^2}{2}+A_z\cos(k_y y)\right)^2A_z$$
$$-\eta_4\left(-\frac{x^2}{2}+A_z\cos(k_y y)\right)A_z+|A_z|^2A_z=0. \quad (17)$$

Here,

$$\eta_1=\frac{\left(2k^2\omega^2\lambda_e^4+2\omega^2\lambda_e^2-k^2\lambda_i^2v_A^2-\lambda_i^2v_A^2k_z^2\right)}{\lambda_i^2v_A^2k^2},$$
$$\eta_2=\frac{4\omega^2k_\perp^2\lambda_e^2}{\lambda_i^2v_A^2k^2k_z^2},\quad \eta_3=\frac{k_y^2}{k_z^2},\quad \eta_4=2\frac{k_y}{k_z}.$$

The expression for A_z in relation to the eikonal[47] S can be formulated as follows: $A_z=A_0(r,z)e^{ik_z(z)S(r,z)}$, where A_z and S are the real functions dependent on r, z, and $r=\sqrt{(x^2+y^2)}$. When we substitute the proposed solution for A_z into Eq. (17) and equate the real and imaginary terms, we obtain the following equations:

Real part:

$$-2\eta_1A_0\frac{\partial S}{\partial z}+\eta_2\lambda_e^2\left(\frac{\partial^2A_0}{\partial x^2}+\frac{\partial^2A_0}{\partial y^2}\right)$$
$$-\eta_2\lambda_e^2k_z^2A_0\left[\left(\frac{\partial S}{\partial x}\right)^2+\left(\frac{\partial S}{\partial y}\right)^2\right]-\eta_3\left(-\frac{x^2}{2}+A_z\cos(k_y y)\right)^2A_0$$
$$-\eta_4\left(-\frac{x^2}{2}+A_z\cos(k_y y)\right)A_0+|A_0|^2A_0=0. \quad (18)$$

Imaginary part:

21 February 2024 06:12:46

Published under an exclusive license by AIP Publishing

$$\frac{2\eta_1}{k_z}\frac{\partial A_0}{\partial z}+2k_z\eta_2\lambda_e^2\frac{\partial A_0}{\partial x}\frac{\partial S}{\partial x}+k_z\eta_2\lambda_e^2A_0\frac{\partial^2 S}{\partial x^2}+2k_z\eta_2\lambda_e^2\frac{\partial A_0}{\partial y}\frac{\partial S}{\partial y}+k_z\eta_2\lambda_e^2A_0\frac{\partial^2 S}{\partial y^2}=0. \tag{19}$$

The initial intensity distribution of the whistler wave is assumed to be Gaussian, i.e., $A_0^2=\frac{A_0'^2}{f_1(z)f_2(z)}\exp\left(-\frac{x^2}{r_{01}^2f_1^2}-\frac{y^2}{r_{02}^2f_2^2}\right)$, and S is represented as $S=\beta_1(z)\frac{x^2}{2}+\beta_2(z)\frac{y^2}{2}$. Here, r_{01} and r_{02} denote the transverse scale size of the whistler wave, while S is a slowly varying function concerning x, y, and z. The functions $f_1(z)$ and $f_2(z)$ serve as a dimensionless parameter, indicating the wave's beam width, $[\beta_1(z)]^{-1}$ and $[\beta_2(z)]^{-1}$ characterize the radius of curvature of the wavefront. Substituting the expressions for A_0^2 and S into Eq. (19), we obtain the following equation:

$$-\frac{2\eta_1}{k_z}\left[\frac{1}{f_1}\frac{df_1}{dz}+\frac{1}{f_2}\frac{df_2}{dz}-\frac{2x^2}{r_{01}^2f_1^3}\frac{df_1}{dz}-\frac{2y^2}{r_{02}^2f_2^3}\frac{df_2}{dz}\right]-2k_z\lambda_e^2\eta_2\left(\frac{2x^2\beta_1}{r_{01}^2f_1^2}+\frac{2y^2\beta_2}{r_{02}^2f_2^2}\right)+k_z\lambda_e^2\eta_2\beta_1(z)+k_z\lambda_e^2\eta_2\beta_2(z)=0. \tag{20}$$

Furthermore, equating the coefficient of x^2 and y^2, we obtain the following relation between wave's beam width parameter $\beta_1(z)$ and $f_1(z)$. Similarly, we get the relation between wave's beam widths $\beta_2(z)$ and $f_2(z)$

$$\beta_1(z)=\frac{\eta_1}{\eta_2\lambda_e^2k_z^2}\frac{1}{f_1}\frac{df_1}{dz},\quad\text{and}\quad\beta_2(z)=\frac{\eta_1}{\eta_2\lambda_e^2k_z^2}\frac{1}{f_2}\frac{df_2}{dz}.$$

Next, by substituting the previously mentioned derivatives of A_0^2 and S parameters into Eq. (18), which represents the real part of Eq. (17)

$$-2\eta_1\frac{x^2}{2}A_0\frac{\partial\beta_1}{\partial z}-2\eta_1\frac{y^2}{2}A_0\frac{\partial\beta_2}{\partial z}+\frac{\eta_2\lambda_e^2}{2}A_0\left[\left(\frac{-2}{r_{01}^2f_1^2}\right)+\left(\frac{-2x}{r_{01}^2f_1^2}\right)^2+\left(\frac{-2}{r_{02}^2f_2^2}\right)+\left(\frac{-2y}{r_{02}^2f_2^2}\right)^2\right]-k_z\lambda_e^2\eta_2A_0(\beta_1(z)\,x)^2-k_z\lambda_e^2\eta_2A_0(\beta_2(z)\,y)^2-\eta_3k_y^2y^2|A_0|^2A_0-\eta_4\frac{x^2}{\lambda_e^2}A_0-\frac{x^2\alpha A_0^2}{r_{01}^2f_1^3f_2}\exp\left(\frac{\alpha A_0^2}{f_1f_2}\right)A_0-\frac{y^2\alpha A_0^2}{r_{02}^2f_2^3f_1}\exp\left(\frac{\alpha A_0^2}{f_1f_2}\right)A_0=0. \tag{21}$$

Once more, compare the coefficients of x^2 and y^2 in Eq. (21).

Coefficients of x^2:

$$-\eta_1\frac{\partial\beta_1}{\partial z}+2\eta_2\lambda_e^2\left(\frac{1}{r_{01}^2f_1^2}\right)^2-k_z\lambda_e^2\eta_2(\beta_1(z)\,)^2-\eta_4\frac{1}{\lambda_e^2}-\frac{\alpha A_0^2}{r_{01}^2f_1^3f_2}\exp\left(\frac{\alpha A_0^2}{f_1f_2}\right)=0. \tag{22}$$

Coefficients of y^2:

$$-\eta_1\frac{\partial\beta_2}{\partial z}+2\eta_2\lambda_e^2\left(\frac{1}{r_{02}^2f_2^2}\right)^2-k_z\lambda_e^2\eta_2(\beta_2(z))^2-\eta_3k_y^2(\alpha A_0^2)A_0-\frac{\alpha A_0^2}{r_{02}^2f_2^3f_1}\exp\left(\frac{\alpha A_0^2}{f_1f_2}\right)A_0=0. \tag{23}$$

Here, $\frac{\partial\beta_1(z)}{\partial z}=\frac{\eta_1}{\eta_2\lambda_e^2k_z^2}\left[-\frac{1}{f_1^2}\left(\frac{df_1}{dz}\right)^2+\frac{1}{f_1}\frac{d^2f_1}{dz^2}\right]$, and $\frac{\partial\beta_2(z)}{\partial z}=\frac{\eta_1}{\eta_2\lambda_e^2k_z^2}\left[-\frac{1}{f_2^2}\left(\frac{df_2}{dz}\right)^2+\frac{1}{f_2}\frac{d^2f_2}{dz^2}\right]$.

By substituting the derivatives of $\beta_1(z)$ and $\beta_2(z)$ into Eqs. (22) and (23), while employing the paraxial approximation $\left(x\ll r_{01}f_1^2,\ y\ll r_{02}f_2^2\right)$, we obtain the equations that govern the beam widths f_1 and f_2 in the x and y directions, respectively,

$$\frac{d^2f_1}{dz^2}=\frac{2\eta_2^2\lambda_e^4k_z^2}{r_{01}^4\eta_1^2f_1^3}-\frac{\eta_2\eta_4k_z^2}{\eta_1^2}f_1-\frac{\lambda_e^2k_z^2\eta_2(\alpha A_0^2)}{r_{01}^2\eta_1^2f_1^2f_2}\exp\left(\frac{\alpha A_0^2}{f_1f_2}\right) \tag{24}$$

and

$$\frac{d^2f_2}{dz^2}=\frac{2\eta_2^2\lambda_e^4k_z^2}{r_{02}^4\eta_1^2f_2^3}-\frac{\eta_2\eta_3k_z^2k_y^2\lambda_e^2(\alpha A_0^2)}{\eta_1^2}f_2-\frac{\lambda_e^2k_z^2\eta_2(\alpha A_0^2)}{r_{02}^2\eta_1^2f_2^2f_1}\exp\left(\frac{\alpha A_0^2}{f_1f_2}\right). \tag{25}$$

Now, we can drive an equation for the dimensionless parameters f_1 and f_2 by utilizing the normalizing parameter $\xi=z/R_d$, here $R_d=k_zr_{01}^2$

$$\frac{d^2f_1}{d\xi^2}=\frac{2\eta_2^2\lambda_e^4k_z^4}{\eta_1^2}\frac{1}{f_1^3}-\frac{\eta_2\eta_4k_z^4r_{01}^4}{\eta_1^2}f_1-\frac{\lambda_e^2k_z^4\eta_2r_{01}^2(\alpha A_0^2)}{\eta_1^2f_1^2f_2}\exp\left(\frac{\alpha A_0^2}{f_1f_2}\right), \tag{26}$$

$$\frac{d^2f_2}{dz^2}=\frac{2\eta_2^2\lambda_e^4k_z^4}{\eta_1^2}\frac{1}{f_2^3}-\frac{\eta_2\eta_3k_z^4k_y^2\lambda_e^2r_{02}^4}{\eta_1^2}(\alpha A_0^2)f_2-\frac{\lambda_e^2k_z^4\eta_2r_{02}^2(\alpha A_0^2)}{\eta_1^2f_2^2f_1}\exp\left(\frac{\alpha A_0^2}{f_1f_2}\right). \tag{27}$$

In Eqs. (26) and (27), the first term accounts for the spreading effect, while the second term and the last term contribute to the convergence of structures. These effects are primarily driven by the presence of the preexisting magnetic island and the ponderomotive nonlinearity. The localization of the whistler wave occurs as a consequence of the competition between the spreading effect term and the last two terms of Eqs. (26) and (27). It is noted in Eq. (27) that the second term involves the whistler wave power. In this case, the whistler wave power affects the current sheet scale size along the y-axis. These f_1 and f_2 equations exhibit anisotropic behavior along the x and y directions, resulting in different rates of change for the dimensionless beam width parameters f_1 and f_2. To solve Eqs. (26) and (27) for an initial plane wavefront, we assume the following initial conditions $f_1=f_2=1$ at $\xi=0$ and $df_1/d\eta=0$, $df_2/d\eta=0$ at $\xi=0$. We have solved our dynamical equation using the semianalytical way in order to comprehend the physical process of coherent structures/current sheet formation and to calculate the scale size of the structures. The semi-analytical study reveals that for a typical set of parameters, when the whistler power $(\alpha A_0^2)=1$, the critical scale size of the localized structures of whistler waves comes out to be $r_{01}=5\lambda_e$ and $r_{02}=2\lambda_e$. The current sheet's scale size can be analyzed using Ampère's law, where the current density is expressed as $j=-\nabla^2A_0$. One can calculate the current sheet transverse scale size comes out to be approximately $r_{01}/\sqrt{3}$ and $r_{02}/\sqrt{3}$. Consequently, the current sheet exhibits a scale size of approximately $2\lambda_e$ along the x-direction and approximately λ_e along the

21 February 2024 06:12:46

Published under an exclusive license by AIP Publishing

y-direction. Recent research by Burch *et al.*[48] has revealed that whistler waves exhibit localized structures with scale sizes ranging from $15\lambda e$ to $20\lambda e$ in magnetopause reconnection sites. Leonenko *et al.*[49] have reported the scale size of superthin electron scale current sheets $0.3\lambda e$ to $1\lambda e$ in the Earth's magnetotail.

V. RESULT AND DISCUSSIONS

Figure 1 depicts the localized structure and its evolution. Whistler waves propagate in the presence of reconnection sites and perturbed density. The background density gets perturbed due to presence of the ion acoustic wave in the background medium. As we observed at an initial time, one can see localization in spatial scales changes slowly itself. Further, these field localized structures have been broken into new localized structures with different intensities. These are consequences of the nonlinear interaction between a 3D whistler wave and a preexisting reconnection site. The whistler wave induces modifications in both the plasma density and the magnetic field, subsequently influencing its own propagation dynamics. As whistler waves traverse through this perturbed medium, they initiate the evolution of a field within the x–y plane and its subsequent amplification. These structures represent the turbulence generation, which provides a passage in the context of energy transfer mechanisms, heating, and particle acceleration within the plasma. Chaston *et al.*[43] proposed wave mode physics using cluster data to explain the power law in the magnetic reconnection site of magnetopause. The coherent structures can take the form of filaments with scale sizes varying from ion scales to electron scales and that the emergence of coherent structures leads to the generation of turbulence. These nonlinear structures represent abrupt phenomena that play a pivotal role in driving the energy transfer process within plasma. These structures serve as catalysts for turbulence and facilitate the transfer of energy from larger spatial scales to smaller ones within reconnection sites.[50–53]

Figure 2 illustrates the appearance of current sheet structures at electron inertial length scales, which have been examined through numerical analysis. These current sheets transfer the injected energy from magnetohydrodynamic (MHD) scales to smaller scales. The formation of spatially and temporally chaotic structures arises as a consequence of nonlinearity. These chaotic configurations have been proposed and substantiated in the context of turbulence within reconnection sites situated at the dayside magnetopause. Recent investigations are going on to understand more about critical roles played by current sheets in particle acceleration and heating processes.[54–56] The present model explores the role of whistler waves in turbulence generation by taking the powerful whistler wave. Here, whistler wave is powerful enough to trigger ponderomotive effects and modification in background density. This includes background density fluctuations arising from the ponderomotive nonlinearity of the wave and the magnetic field perturbation due to the presence of preexisting magnetic islands. These results represent that considering the field fluctuations resulting from the presence of magnetic islands, where the coefficient of the perturbing amplitude is whistler wave field, have an impact on the evolution of the current sheets. Whistler waves significantly alter the aspect ratio of the current sheet during a magnetopause reconnection event (at electron diffusion region).[15,57,58]

An ensemble averaged wave number spectrum of the magnetic field is presented in Fig. 3. The blue line is the scaling reference line in the observed spectrum that Chaston *et al.*[43] reported. This power spectrum follows Kolmogorov's $k^{-3/2}$ in the inertial scale and further gets steeper with k^{-3} scaling. In Fig. 3, the power spectrum has to steepen at small scales. This steepening exhibits energy cascading phenomena.

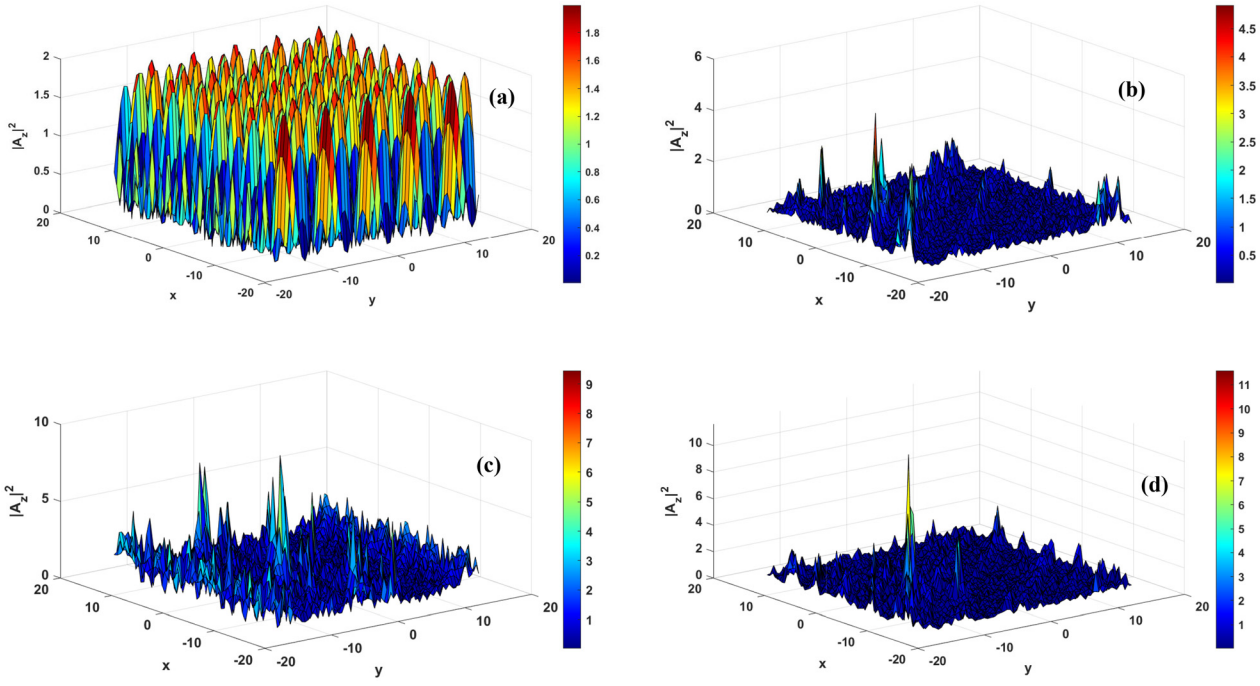


FIG. 1. The spatial distribution of the vector potential in the $x - y$ plane at (a) $t = 0$, (b) $t = 40$, (c) $t = 50$, and (d) $t = 60$.

21 February 2024 06:12:46

Published under an exclusive license by AIP Publishing

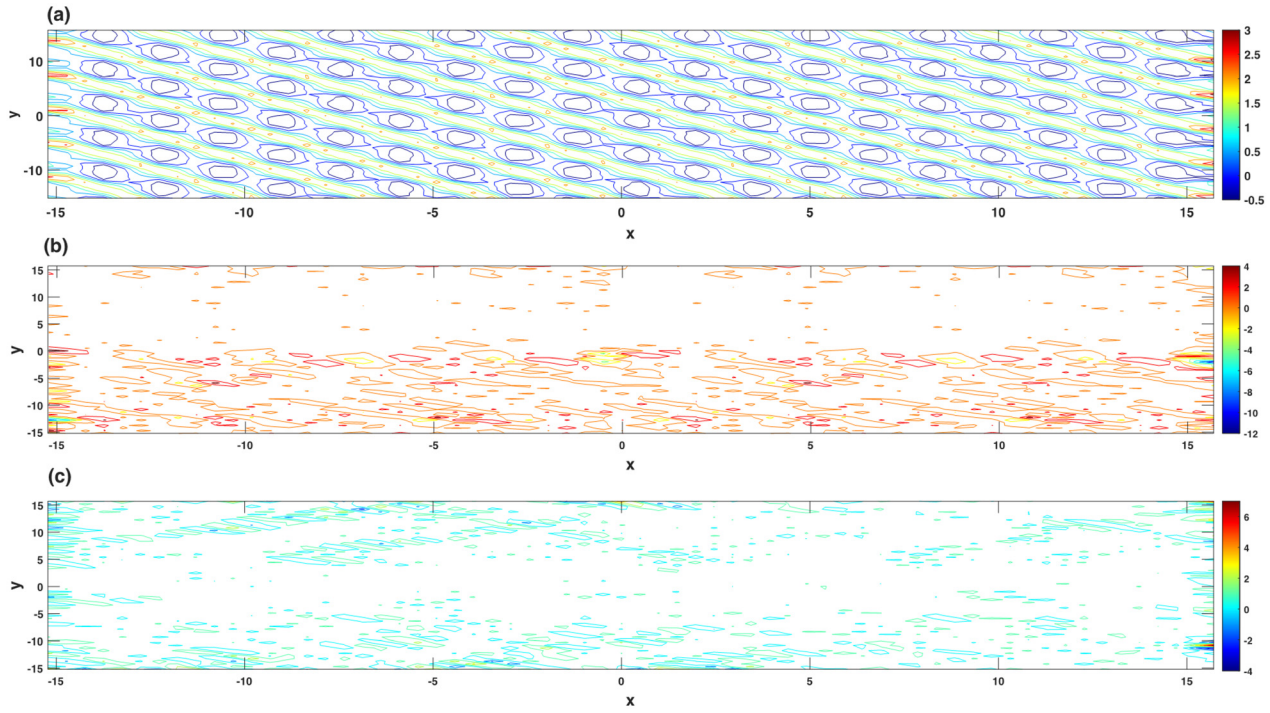


FIG. 2. Plot of magnetic flux at different times (a) $t = 0$, (b) $t = 4$, and (c) $t = 17$.

The inertial range scaling $k^{-3/2}$ is consistent with the classical Kolmogorov slope in the inertial range for turbulence. The magnetic field power spectrum has a steeper slope of k^{-3}, which is observed by Chaston *et al.*[43] at Earth's magnetopause.[59]

VI. SUMMARY

In previous work, Pathak *et al.*[40] considered field fluctuations resulting from the presence of magnetic islands, where the coefficient of the perturbing amplitude is constant. However, in the present model, whistler is powerful enough that could impact magnetic islands. In the present model, we consider the whistler wave is powerful and we take into account the field fluctuations resulting from the presence of magnetic islands, where the whistler wave field could be the coefficient of the perturbing amplitude. Also, whistler wave is powerful enough to trigger ponderomotive effects and modification in background density. To investigate the generation of turbulence associated with whistler waves, we have developed the nonlinear dynamical equation representing the propagation and interaction with a magnetic island near the dayside magnetosphere reconnection site and performed a numerical simulation. The present manuscript deals with spatial and temporal evolution of whistler waves. For this study, the dynamical equations are derived by considering the ponderomotive

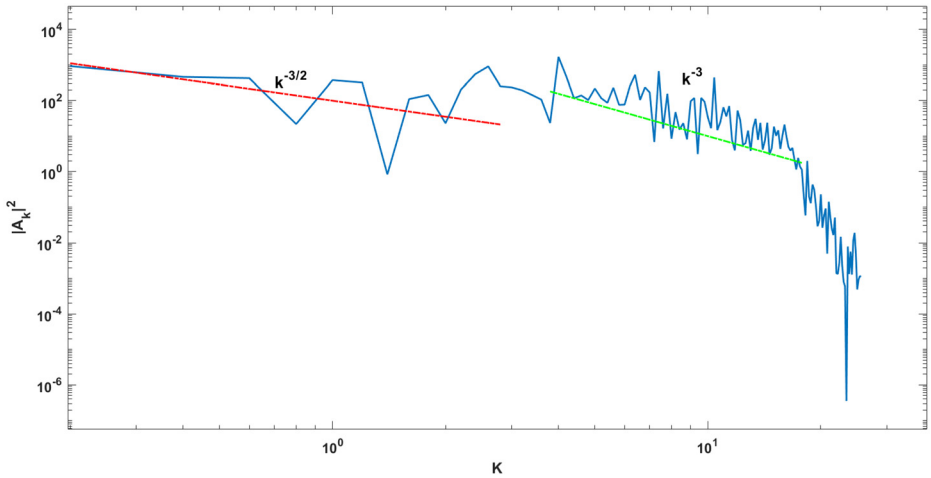


FIG. 3. The magnetic power spectrum associated with whistler waves at magnetopause reconnection sites.

21 February 2024 06:12:46

Published under an exclusive license by AIP Publishing

force-driven density modification and magnetic field perturbation due to the shear field modeled by the magnetic island. Further, numerical simulation is done using the numerical method (the finite difference method for time scales and the spectral method for spatial scales).

Using typical parameters[43] for plasma in the magnetopause region, the numerical simulation is processed. As shown in the results, the magnetic island and the density modification affect the generation of coherent structures and support in the generation of turbulence. Under the paraxial approximation, a simple analytical model is used to study the critical size of localized structures. Using the magnetospheric multiscale (MMS) mission, whistler mode waves have been observed in the magnetic reconnection zone close to the dayside magnetopause.[60] Whistler waves are most often seen on the magnetospheric side of the center plane during this occurrence of reconnection events. This research is more useful for the MMS mission observation, focusing on electron-scale structures during reconnection. Zhao *et al.*[61] presented dynamics of charged particles and whistler waves in Earth's magnetosheath. Our present model explores the role of whistler wave in the presence of the ponderomotive force-driven density modification and magnetic field perturbation due to preexisting magnetic islands. Additionally, these magnetic islands are themselves impacted by wave power considered. The uncertainty in determining the thickness of the embedded superthin electron current sheets has persisted due to the scale-free nature of the motion of magnetized electrons. Zelenyi *et al.*[62] proposed an analytical model that establishes a spatial scaling for the inner (embedded) electron-scale current layer, dependent on the local characteristics of the thin current sheets, i.e., proton plasma frequency, gyroradius of thermal protons. Leonenko *et al.*[49] observed superthin electron-scale current sheets, which have spatial scaling $\sim 0.3\,\lambda_e - 1\,\lambda_e$. They found that in some cases, evolution of the superthin electron-scale current sheets characterized by unmagnetized electrons and it is similar to the mechanism of ion-scale thin current sheet formation by the quasi-adiabatic ions.[63] In the present model, we calculate the scale size of the current sheet approximately $2\lambda_e$ along the x-direction and approximately λ_e along the y-direction. From this model, we have been able to elucidate the formation of superthin electron current sheets at electron scales, providing valuable insights into the role of a whistler wave power to the evolution of current sheets at electron scales.

ACKNOWLEDGMENTS

This work was supported by the Scientific Engineering Research Board (SERB), Department of Science and Technology, Government of India, under Grant No. PDF/2019/000891.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Neha Pathak: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Methodology (lead); Software (lead); Validation (lead); Visualization (lead); Writing – original draft (lead). **Suresh C. Sharma:** Conceptualization (equal); Resources (equal); Supervision (equal). **Garima Patel:** Formal analysis (equal); Writing – review & editing (equal). **Rampal Sharma:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Supervision (equal); Visualization (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article.

REFERENCES

[1]S. R. Cranmer, M. Asgari-Targhi, M. P. Miralles, J. C. Raymond, L. Strachan, H. Tian, and L. N. Woolsey, "The role of turbulence in coronal heating and solar wind expansion," Philos. Trans. R. Soc., A **373**(2041), 20140148 (2015).
[2]W. H. Matthaeus, J. W. Bieber, and G. P. Zank, "Unquiet on any front: Anisotropic turbulence in the solar wind," Rev. Geophys. **33**(1S), 609, https://doi.org/10.1029/95RG00496 (1995).
[3]J. E. Borovsky, R. C. Elphic, H. O. Funsten, and M. F. Thomsen, "The Earth's plasma sheet as a laboratory for flow turbulence in high-β MHD," J. Plasma Phys. **57**(1), 1 (1997).
[4]A. Chasapis, W. H. Matthaeus, T. N. Parashar, O. LeContel, A. Retinò, H. Breuillard, Y. Khotyaintsev, A. Vaivads, B. Lavraud, E. Eriksson, T. E. Moore, J. L. Burch, R. B. Torbert, P.-A. Lindqvist, R. E. Ergun, G. Marklund, K. A. Goodrich, F. D. Wilder, M. Chutter, J. Needell, D. Rau, I. Dors, C. T. Russell, G. Le, W. Magnes, R. J. Strangeway, K. R. Bromund, H. K. Leinweber, F. Plaschke, D. Fischer, B. J. Anderson, C. J. Pollock, B. L. Giles, W. R. Paterson, J. Dorelli, D. J. Gershman, L. Avanov, and Y. Saito, "Electron heating at kinetic scales in magnetosheath turbulence," Astrophys. J. **836**(2), 247 (2017).
[5]J. E. Stawarz, R. E. Ergun, and K. A. Goodrich, "Generation of high-frequency electric field activity by turbulence in the Earth's magnetotail," J. Geophys. Res. Space Phys. **120**(3), 1845, https://doi.org/10.1002/2014JA020166 (2015).
[6]J. E. Stawarz, S. Eriksson, F. D. Wilder, R. E. Ergun, S. J. Schwartz, A. Pouquet, J. L. Burch, B. L. Giles, Y. Khotyaintsev, O. L. Contel, P. A. Lindqvist, W. Magnes, C. J. Pollock, C. T. Russell, R. J. Strangeway, R. B. Torbert, L. A. Avanov, J. C. Dorelli, J. P. Eastwood, D. J. Gershman, K. A. Goodrich, D. M. Malaspina, G. T. Marklund, L. Mirioni, and A. P. Sturner, "Observations of turbulence in a Kelvin–Helmholtz event on 8 September 2015 by the Magnetospheric Multiscale mission," J. Geophys. Res. Space Phys. **121**(11), 11021 (2016).
[7]Z. Vörös, W. Baumjohann, R. Nakamura, M. Volwerk, and A. Runov, "Bursty bulk flow driven turbulence in the earth's plasma sheet," Space Sci. Rev. **122**, 301 (2006).
[8]G. Zimbardo, A. Greco, L. Sorriso-Valvo, S. Perri, Z. Vörös, G. Aburjania, K. Chargazia, and O. Alexandrova, "Magnetic turbulence in the geospace environment," Space Sci. Rev. **156**(1–4), 89 (2010).
[9]J. F. Drake, M. Swisdak, H. Che, and M. A. Shay, "Electron acceleration from contracting magnetic islands during reconnection," Nature **443**(7111), 553 (2006).
[10]J. P. Eastwood, T. D. Phan, S. D. Bale, and A. Tjulin, "Observations of turbulence generated by magnetic reconnection," Phys. Rev. Lett. **102**(3), 035001 (2009).
[11]W. Daughton, T. K. M. Nakamura, H. Karimabadi, V. Roytershteyn, and B. Loring, "Computing the reconnection rate in turbulent kinetic layers by using electron mixing to identify topology," Phys Plasmas **21**(5), 052307 (2014).
[12]J. T. Gosling, "Observations of magnetic reconnection in the turbulent high-speed solar wind," Astrophys. J. **671**(1), L73 (2007).
[13]A. Retinò, D. Sundkvist, A. Vaivads, F. Mozer, M. André, and C. J. Owen, "*In situ* evidence of magnetic reconnection in turbulent plasma," Nat. Phys. **3**(4), 235 (2007).
[14]R. E. Ergun, N. Ahmadi, L. Kromyda, S. J. Schwartz, A. Chasapis, S. Hoilijoki, F. D. Wilder, J. E. Stawarz, K. A. Goodrich, D. L. Turner, I. J. Cohen, S. T. Bingham, J. C. Holmes, R. Nakamura, F. Pucci, R. B. Torbert, J. L. Burch, P.-A. Lindqvist, R. J. Strangeway, O. Le Contel, and B. L. Giles, "Observations of particle acceleration in magnetic reconnection–driven turbulence," Astrophys. J. **898**(2), 154 (2020).

21 February 2024 06:12:46

Published under an exclusive license by AIP Publishing

[15]Y. V. Khotyaintsev, D. B. Graham, C. Norgren, and A. Vaivads, "Collisionless magnetic reconnection and waves: Progress review," Front. Astron. Space Sci. **6**, 70 (2019).
[16]H. Viberg, Y. V. Khotyaintsev, A. Vaivads, M. André, and J. S. Pickett, "Mapping HF waves in the reconnection diffusion region," Geophys. Res. Lett. **40**(6), 1032, https://doi.org/10.1002/grl.50227 (2013).
[17]M. Zhou, X. H. Deng, S. Y. Li, Y. Pang, A. Vaivads, H. Réme, E. Lucek, S. Fu, X. Lin, Z. G. Yuan, and J. F. Wang, "Observation of waves near lower hybrid frequency in the reconnection region with thin current sheet," J. Geophys. Res. Space Phys. **114**(2), A02216 (2009).
[18]F. D. Wilder, S. J. Schwartz, R. E. Ergun, S. Eriksson, N. Ahmadi, A. Chasapis, D. L. Newman, J. L. Burch, R. B. Torbert, R. J. Strangeway, and B. L. Giles, "Parallel electrostatic waves associated with turbulent plasma mixing in the Kelvin–Helmholtz instability," Geophys. Res. Lett. **47**(14), e2020GL087837, https://doi.org/10.1029/2020GL087837 (2020).
[19]X. J. Zhang, L. Chen, A. V. Artemyev, V. Angelopoulos, and X. Liu, "Periodic excitation of chorus and ECH waves modulated by ultralow frequency compressions," J. Geophys. Res. Space Phys. **124**(11), 8535, https://doi.org/10.1029/2019JA027201 (2019).
[20]N. Pathak, R. Uma, and R. P. Sharma, "Localized structures and turbulent spectra in the magnetopause," J. Geophys. Res. Space Phys. **124**(4), 2515, https://doi.org/10.1029/2018JA026147 (2019).
[21]X. An, J. Li, J. Bortnik, V. Decyk, C. Kletzing, and G. Hospodarsky, "Unified view of nonlinear wave structures associated with whistler-mode chorus," Phys. Rev. Lett. **122**(4), 045101 (2019).
[22]J. Ambrosiano, W. H. Matthaeus, M. L. Goldstein, and D. Plante, "Test particle acceleration in turbulent reconnecting magnetic fields," J. Geophys. Res. Space Phys. **93**(A12), 14383, https://doi.org/10.1029/JA093iA12p14383 (1988).
[23]Z. Z. Chen, J. Yu, J. Wang, Z. G. He, N. G. Liu, J. Cui, and J. B. Cao, "High-frequency electrostatic waves modulated by whistler waves behind dipolarization front," J. Geophys. Res. Space Phys. **127**(11), e2022JA030935, https://doi.org/10.1029/2022JA030935 (2022).
[24]Z. Guo, L. Dai, Y. Ren, and C. Wang, "MMS observations of lower hybrid and whistler waves inside a foreshock transient," J. Geophys. Res. Space Phys. **127**(11), e2022JA030707, https://doi.org/10.1029/2022JA030707 (2022).
[25]S. Y. Huang, F. Sahraoui, Z. G. Yuan, O. L. Contel, H. Breuillard, J. S. He, J. S. Zhao, H. S. Fu, M. Zhou, X. H. Deng, X. Y. Wang, J. W. Du, X. D. Yu, D. D. Wang, C. J. Pollock, R. B. Torbert, and J. L. Burch, "Observations of whistler waves correlated with electron-scale coherent structures in the magnetosheath turbulent plasma," Astrophys. J. **861**(1), 29 (2018).
[26]R. L. Stenzel, "Whistler waves in space and laboratory plasmas," J. Geophys. Res. Space Phys. **104**(A7), 14379, https://doi.org/10.1029/1998JA900120 (1999).
[27]D. Biskamp, E. Schwarz, A. Zeiler, A. Celani, and J. F. Drake, "Electron magnetohydrodynamic turbulence," Phys. Plasmas **6**(2–3), 751 (1999).
[28]S. Dastgeer and G. P. Zank, "Anisotropic turbulence in two-dimensional electron magnetohydrodynamics," Astrophys. J. **599**(1), 120502 (2003).
[29]S. Galtier and A. Bhattacharjee, "Anisotropic weak whistler wave turbulence in electron magnetohydrodynamics," Phys. Plasmas **10**(8), 3065 (2003).
[30]J. Cho and A. Lazarian, "Simulations of electron magnetohydrodynamic turbulence," Astrophys. J. **701**(1), 236 (2009).
[31]J. S. Zhao, J. Y. Lu, and D. J. Wu, "Parametric instability of whistler waves in the electron magnetohydrodynamics," Astrophys. J. **714**(1), 138 (2010).
[32]J. Zhao, "Properties of whistler waves in warm electron plasmas," Astrophys. J. **850**(1), 13 (2017).
[33]W. H. Matthaeus and M. L. Goldstein, "Measurement of the rugged invariants of magnetohydrodynamic turbulence in the solar wind," J. Geophys. Res. Space Phys. **87**(A8), 6011, https://doi.org/10.1029/JA087iA08p06011 (1982).
[34]A. A. Schekochihin, S. C. Cowley, W. Dorland, G. W. Hammett, G. G. Howes, E. Quataert, and T. Tatsuno, "Astrophysical gyrokinetics: Kinetic and fluid turbulent cascades in magnetized weakly collisional plasmas," Astrophys. J., Suppl. Ser. **182**(1), 130 (2009).
[35]D. Biskamp, E. Schwarz, and J. F. Drake, "Two-fluid theory of collisionless magnetic reconnection," Phys. Plasmas **4**(4), 1002 (1997).
[36]R. P. Sharma, N. Pathak, N. Yadav, and P. Sharma, "Nonlinear propagation of whistler wave and turbulent spectrum in reconnection region of magnetopause," Phys. Plasmas **24**(9), 092902 (2017).
[37]R. Fitzpatrick and F. L. Waelbroeck, "Two-fluid magnetic island dynamics in slab geometry. II. Islands interacting with resistive walls or resonant magnetic perturbations," Phys. Plasmas **12**(2), 022308 (2005).
[38]N. Yadav and R. P. Sharma, "Nonlinear interaction of 3D kinetic Alfvén waves and ion acoustic waves in solar wind plasmas," Sol. Phys. **289**(5), 1803 (2014).
[39]N. Pathak, N. Yadav, and R. P. Sharma, "Nonlinear evolution of 3D whistler waves in space plasmas," Phys. Plasmas **24**(6), 062104 (2017).
[40]N. Pathak, N. Yadav, S. Sharma, P. Sharma, and R. P. Sharma, "Localization of whistler wave and turbulent spectra in the magnetotail region," J. Geophys. Res. Space Phys. **122**(2), 1751, https://doi.org/10.1002/2016JA023640 (2017).
[41]R. K. Rai, N. Pathak, P. Sharma, S. Sharma, N. Yadav, and R. P. Sharma, "Turbulence generation of ion scale in the presence of magnetic islands and guide field at the magnetopause region," J. Astrophys. Astron. **42**(1), 1 (2021).
[42]H. Dewan, N. Pathak, R. Uma, and R. P. Sharma, "Numerical simulations to study turbulent magnetic field amplification by nonlinear interaction of high-power laser and kinetic Alfvén waves in laboratory and astrophysical plasmas," Phys. Plasmas **26**(12), 122105 (2019).
[43]C. Chaston, J. Bonnell, J. P. McFadden, C. W. Carlson, C. Cully, O. L. Contel, A. Roux, H. U. Auster, K. H. Glassmeier, V. Angelopoulos, and C. T. Russell, "Turbulent heating and cross-field transport near the magnetopause from THEMIS," Geophys. Res. Lett. **35**(17), L17S08 (2008).
[44]N. Gaur and R. P. Sharma, "Nonlinear effects associated with kinetic Alfvén wave," J. Geophys. Res. Space Phys. **120**(4), 2397, https://doi.org/10.1002/2014JA020771 (2015).
[45]S. A. Orszag and C. M. Tang, "Small-scale structure of two-dimensional magnetohydrodynamic turbulence," J. Fluid Mech. **90**(1), 129 (1979).
[46]D. Biskamp and H. Welter, "Dynamics of decaying two-dimensional magnetohydrodynamic turbulence," Phys. Fluids B **1**(10), 1964 (1989).
[47]S. A. Akhmanov, A. P. Sukhorukov, and R. V. Khokhlov, "Self-focusing and diffraction of light in a Nonlinear medium," Usp. Fiz. Nauk **93**(9), 19 (1967).
[48]J. L. Burch, R. E. Ergun, P. A. Cassak, J. M. Webster, R. B. Torbert, B. L. Giles, J. C. Dorelli, A. C. Rager, K. J. Hwang, T. D. Phan, K. J. Genestreti, R. C. Allen, L. J. Chen, S. Wang, D. Gershman, O. L. Contel, C. T. Russell, R. J. Strangeway, F. D. Wilder, D. B. Graham, M. Hesse, J. F. Drake, M. Swisdak, L. M. Price, M. A. Shay, P. A. Lindqvist, C. J. Pollock, R. E. Denton, and D. L. Newman, "Localized oscillatory energy conversion in magnetopause reconnection," Geophys. Res. Lett. **45**(3), 1237, https://doi.org/10.1002/2017GL076809 (2018).
[49]M. V. Leonenko, E. E. Grigorenko, L. M. Zelenyi, H. V. Malova, A. Y. Malykhin, V. Y. Popov, and J. Büchner, "MMS observations of super thin electron-scale current sheets in the earth's magnetotail," J. Geophys. Res. Space Phys. **126**(11), e2021JA029641, https://doi.org/10.1029/2021JA029641 (2021).
[50]W. H. Matthaeus and S. L. Lamkin, "Rapid magnetic reconnection caused by finite amplitude fluctuations," Phys. Fluids **28**(1), 303 (1985).
[51]H. Karimabadi, V. Roytershteyn, M. Wan, W. H. Matthaeus, W. Daughton, P. Wu, M. Shay, B. Loring, J. Borovsky, E. Leonardis, S. C. Chapman, and T. K. M. Nakamura, "Coherent structures, intermittent turbulence, and dissipation in high-temperature plasmas," Phys. Plasmas **20**(1), 012303 (2013).
[52]H. Karimabadi and A. Lazarian, "Magnetic reconnection in the presence of externally driven and self-generated turbulence," Phys. Plasmas **20**(11), 112102 (2013).
[53]H. Karimabadi, V. Roytershteyn, W. Daughton, and Y. H. Liu, "Recent evolution in the theory of magnetic reconnection and its connection with turbulence," Space Sci. Rev. **178**(2–4), 307 (2013).
[54]V. Angelopoulos, J. P. McFadden, D. Larson, C. W. Carlson, S. B. Mende, H. Frey, T. Phan, D. G. Sibeck, K. H. Glassmeier, U. Auster, E. Donovan, I. R. Mann, I. J. Rae, C. T. Russell, A. Runov, X. Z. Zhou, and L. Kepko, "Tail reconnection triggering substorm onset," Science **321**(5891), 931 (2008).
[55]A. T. Y. Lui, M. Vorwerk, M. W. Dunlop, I. V. Alexeev, A. N. Fazakerley, A. P. Walsh, M. Lester, A. Grocott, C. Mouikis, M. G. Henderson, L. M. Kistler, C. Shen, J. K. Shi, T. L. Zhang, and H. Rème, "Near-Earth substorm features from multiple satellite observations," J. Geophys. Res. Space Phys. **113**(7), A07S26 (2008).
[56]V. A. Sergeev, V. Angelopoulos, and R. Nakamura, "Recent advances in understanding substorm dynamics," Geophys. Res. Lett. **39**(5), L05101 (2012).
[57]L. Price, M. Swisdak, J. F. Drake, J. L. Burch, P. A. Cassak, and R. E. Ergun, "Turbulence in three-dimensional simulations of magnetopause reconnection," J. Geophys. Res. Space Phys. **122**(11), 11086 (2017).

21 February 2024 06:12:46

Published under an exclusive license by AIP Publishing

[58]S. Wang, D. B. Graham, X. An, L. Li, Q. G. Zong, X. Z. Zhou, W. Y. Li, and Z. Y. Liu, "Electrostatic waves around a magnetopause reconnection secondary electron diffusion region modulated by whistler and lower-hybrid waves," Geophys. Res. Lett. **50**(18), e2023GL104905, https://doi.org/10.1029/2023GL104905 (2023).
[59]J. L. Burch, T. E. Moore, R. B. Torbert, and B. L. Giles, "Magnetospheric multiscale overview and science objectives," Space Sci. Rev. **199**(1–4), 5 (2016).
[60]F. D. Wilder, R. E. Ergun, K. A. Goodrich, M. V. Goldman, D. L. Newman, D. M. Malaspina, A. N. Jaynes, S. J. Schwartz, K. J. Trattner, J. L. Burch, M. R. Argall, R. B. Torbert, P. A. Lindqvist, G. Marklund, O. L. Contel, L. Mirioni, Y. V. Khotyaintsev, R. J. Strangeway, C. T. Russell, C. J. Pollock, B. L. Giles, F. Plaschke, W. Magnes, S. Eriksson, J. E. Stawarz, A. P. Sturner, and J. C. Holmes, "Observations of whistler mode waves with nonlinear parallel electric fields near the dayside magnetic reconnection separatrix by the Magnetospheric Multiscale mission," Geophys. Res. Lett. **43**(12), 5909, https://doi.org/10.1002/2016GL069473 (2016).
[61]J. Zhao, T. Wang, C. Shi, D. B. Graham, M. W. Dunlop, J. He, B. T. Tsurutani, and D. Wu, "Ion and electron dynamics in the presence of mirror, electromagnetic ion cyclotron, and whistler waves," Astrophys. J. **883**(2), 185 (2019).
[62]L. M. Zelenyi, H. V. Malova, E. E. Grigorenko, V. Y. Popov, and E. M. Dubinin, "Universal scaling of thin current sheets," Geophys. Res. Lett. **47**(14), e2020GL088422, https://doi.org/10.1029/2020GL088422 (2020).
[63]J. Büchner and L. M. Zelenyi, "Regular and chaotic particle motion in sheared magnetic field reversals," Adv. Space Res. **11**(9), 177 (1991).

21 February 2024 06:12:46

Published under an exclusive license by AIP Publishing

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-18331-8

Seagull optimization algorithm for node localization in wireless sensor networks

Yogendra Mohan[1,2] · **Rajesh Kumar Yadav**[2] · **Manisha Manjul**[3]

Received: 29 November 2022 / Revised: 4 January 2024 / Accepted: 19 January 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
Numerous applications of wireless sensor networks (WSNs) highly depend on the node location, such as maritime rescue, agriculture, and hazardous environments. GPS-enabled sensors are not cost-effective or energy efficient. Henceforth, the precise location of sensor nodes significantly impacts the performance of WSNs. Finding the position of a target node (an unknown node) is known as node localization. Gaussian noise and RSSI are cost-effective approaches for estimating the location of a target node. In this paper, the Seagull optimization algorithm and its enhanced versions are applied to increase the NL accuracy of localized nodes. To enhance the accuracy and improve the randomness of the seagull optimization algorithm (SOA)., levy flight and a chaotic map are employed in this work to enhance the seagull optimization algorithm Furthermore, the chaotic map-based SOA (C-SOA) and the Levy flight-based SOA (LF-SOA) are used for node location in WSNs. The performance evaluation and result comparison of SOA, C-SOA, and LF-SOA show that LF-SOA is better than C-SOA and SOA.

Keywords Node localization (NL) · Seagull optimization algorithm (SOA) · Chaotic map · Levy flight

1 Introduction

Wireless sensor nodes are tiny, low-power gadgets with an antenna, an integrated CPU, and a radio that runs on batteries. To monitor and record important data or events, WSNs [20] are made up of numerous sensor nodes outfitted with specialized transducers and radio communications equipment. WSNs have piqued the interest of all researchers due to their application-oriented approach, which identifies the particular services provided by every

✉ Yogendra Mohan
yogendra.mohan@gmail.com; ymo@nerist.ac.in

[1] Department of Computer Science and Engineering, North Eastern Regional Institute of Science and Technology Nirjuli, Itanagar 791109, Arunachal Pradesh, India

[2] Department of Computer Science and Engineering, Delhi Technological University, New Delhi 110 042, India

[3] Computer Science and Application, G.B. Pant DSEU Okhala-1 Campus, New Delhi 110077, India

application. Henceforth, WSNs are becoming progressively popular in terms of applications, services, and choice. Since these sensor networks have begun to become a part of our daily lives, such as smart homes, cities, agriculture, transportation, etc. With all these benefits, WSNs have a lot of limitations, like the adoption of new technologies like cloud, fog with the emerging applications, and computational and communication resources of sensor nodes. Moreover, WSNs face many challenges, like robust routing, securing devices, and tracking device location (node localization). Among all these, node localizations (NL) are one of the prior and demanding challenges. The sensors can be deployed on earth, underground, or underwater. It is difficult to localize the sensors deployed.

The precise localization of sensor nodes substantially impacts the functioning of WSNs. The process of determining the exact locations of the sensor nodes in WSNs is known as positioning or NL [26–28]. Despite being a standard positioning technology, GPS has a high energy consumption [30, 31], making it unsuitable for NL in WSNs. Numerous variables impact the accuracy of NL, and they are the network design, the density of anchor nodes, and the measuring area's geometric shape. To improve the accuracy of NL, a meta-heuristic optimization algorithm is modified and used. The simulation is carried out for sensor networks of different sizes, ranging from 25 to 150 TNs and 15 to 90 ANs, and the distance between ANs and TN is calculated using the RSSI and Gaussian mixture model.

In localization, meta-heuristic optimization algorithms [1, 7–9, 41, 42, 46–48] are gaining popularity. Seagull lives in colonies; the seagull's migration patterns and attacking behaviors inspire the seagull optimization scheme. The migration and attacking behavior of the seagull are mathematically modeled with the exploration and exploitation phase. Various research work has accounted for that exploration is essential for addressing optimization issues. The migration and attacking behavior of seagull birds are modeled to estimate the target nodes. In the exploration phase, Seagulls move and change their position and find the best position. In the exploitation phase, seagulls attack the prey. The existing SOA algorithm is based on a random number. Although SOA has a global exploration potential, it has drawbacks such as late convergence and a tendency to get stuck in local optima. That is why SOA is modified using a chaotic map and levy flight. The comparative performance analysis among the SOA, C-SOA, and LF-SOA is performed, and according to experimental findings, it is confirmed that the LF-SOA significantly outperforms the SOA and C-SOA in terms of NL accuracy and NL time.

The following are the major contributions of the current article:

- The existing SOA algorithms are modified using a chaotic map and levy flight schemes. The modified meta-heuristic algorithms, namely C-SOA and LF-SOA, have been used for NL optimization.
- In C-SOA, ten different chaotic maps have been chosen for the performance test. Out of these, the logistic map was found to be best suitable for exploration.
- In LF-SOA, exploration power is provided by the levy-flight distribution.
- Comparative analysis among SOA, C-SOA, and LF-SOA is performed.

The novelty of this paper is to enhance the accuracy of the NL optimization, SOA hybridized using levy-flight and chaotic map to emphasize on scalability and robustness to node failures. There are several advantages of using chaotic maps and levy-flight for NL optimization as follows. *Scalability*: In large-scale WSNs, NL using chaotic maps is useful. The network adapts to changes in size and density while retaining precise localization thanks to their help in organizing and coordinating

nodes effectively. *Robustness to Node Failures*: Chaotic maps offer resilience to node failures. Chaotic maps can aid other nodes in swiftly adjusting to the changing network architecture and allowing localization to continue without major interruptions when a node fails or departs the network.

The remaining parts of the manuscript are divided into the following sections: Sect. 2 is the related work of different meta-heuristic optimization algorithms dealing with NL. Section 3 briefly introduces the chaotic map, levy flight, and SOA. Section 4 covers the proposed C-SOA and LF-SOA. Section 5 analyzes the NL simulation results of SOA, C-SOA, and LF-SOA. Section 6 presents the conclusion and future research work.

2 Related work

This section reviews the literature work based on node localization for WSNs. Yaming et al. [24] have proposed a scheme that is found suitable for environments where an offline calibration phase for the path-loss component is not feasible. The RSSI technique estimates the distance between the anchor node (AN) and the target node (TN). The simulation results demonstrate that the suggested technique has a better NL accuracy than existing methods. The limitation of the proposed scheme is the same value of the path-loss exponent is taken for the different measurements, which may affect the node localization accuracy. Ziwen et al. [5] have proposed a Multi-objective PSO for the NL. The localization accuracy and average node localization error are improved, while the convergence rate is improved. The performance parameters are convergence rate, node localization accuracy, and average node localization error.

Sankalp et al. [4] have proposed a Butterfly Optimization Algorithm (BFOA) for NL. BFOA is compared with PSO and FFA. The performance of BFOA is better than PSO and FFOA. The trilateration method is used for distance computation between AN and TN, but the value of the environment constant and Path loss component is not mentioned. The performance parameters are the mean of NL, mean of NL error, and mean of computation time. Mihoubi et al. [19] have proposed an RSSI-based Bat optimization algorithm (BOA), and the result obtained shows that the Bat algorithm is more efficient than PSO and DE. RSSI scheme is used for distance estimation between AN and TN, but the value of environment constant and Path loss component is not mentioned. The performance parameters are mean NL error and mean NL time. The simulation result showed that the accuracy and convergence rate of the Bat algorithm is better than PSO and DE. Rajakumar et al. [3] proposed a GWO simulation algorithm. An improved DV-Hop method is proposed to resolve the problems with the DV-Hop algorithm for NL. The performance parameters NL accuracy, computation time, and minimum localization error are compared with PSO and GWO. The convergence rate and NL accuracy are high compared to the PSO and MBA, and NL error is low. Tu Li et al. [11] have proposed a DV-HOP-based PSO algorithm. The chaotic map is hybridized to enhance the NL accuracy and minimize the computation delay. So many chaotic maps are available, and a sine chaotic map is chosen for simulations. It is not mentioned why only a sine map is chosen for it. The performance parameters are NL and NL Error.

Huthaifa et al. [2] have proposed an SSO algorithm for NL. Mean NL error, NL computation time, and the total number of nodes localized are the performance parameters of the proposed scheme. The NL accuracy is improved when the AN density increases concerning

the number of TNs. The computation time for all localization methods is observed to grow as the number of TNs and ANs increases, although the SSOA has the fastest computation time compared to PSO, BOA, FFA, and GWO. The performance of SSOA is better than BOA, FA, PSO, and GWO for all the performance parameters. Ju Liu et al. [37] have proposed a hybrid MCB-PSO algorithm for NL. The WSN environment is 3-dimensional. The anchor and target nodes are mobile. The convergence rate is improved. The main drawback of the MCB method is the random generation of node placements in an anchor box, and there needs to be a searching direction, which results in low localization efficiency and high computation time. The performance parameters are NL error and computation time. Ling Song et al. [36] have proposed the Monte Carlo NL algorithm, which is hybridized with QUasi-Affine TRansformation Evolutionary (QUATRE) optimization. The advantage of the proposed algorithm is that only 30 anchor nodes are deployed for the localization of 320 target nodes, and the enhanced QUATRE optimization-based Monte Carlo NL scheme offers increased positioning accuracy and convergence rate. Although the proposed algorithm requires some hardware, it is cost-effective. Yeasir et al. [35] have proposed a GWO-based compressive sensing (CS-GWO) algorithm. An energy-efficient bio-inspired NL algorithm and clustering algorithms are implemented in Unnamed Aerial Vehicles networks. The limitation of the proposed scheme is all sensor nodes are GPS-enabled, which consumes more power. The performance parameters are Node Localization, Mean Node Localization error, and clustering.

Chuku et al. [25] have proposed utilizing outlier identification techniques to mitigate the impact of such disproportionately inaccurate distance estimates in RSSI-based location estimation. Three distinct node localization methods are proposed to reduce node localization mistakes in shadowed environments that employ outlier identification. NL error is reduced in the proposed scheme. The distance between the AN and TN is estimated using the RSSI of the RF signal, but the environment constant and path loss component value are essential to the accuracy of node localization. The environment constant value lies between 2 and 5, and the fixed value of path loss exponent 2.3 is chosen without any calibration, which may affect the NL accuracy. Pudi et al. [12] have designed a Group Teaching Optimization Algorithm (GTOA) for indoor NL. Comparative results of the GTOA are better than the PSO, krill herd algorithm (KHA), GWO, cuckoo search optimization (CSO), and firefly algorithm (FFA). The performance parameters are NL and NL Error. The KHA, PSO, CSO, GWO, and FFA-based NL algorithms have higher NL errors of 0.36, 0.57, 0.38, 0.53, and 0.51, respectively. The GTOA scheme for NL has achieved a reduced NL error of 0.16. Cheng et al. [18] have proposed a DV-Hop-based Modified Archimedes optimization Algorithm (MAOA) for NL. The convergence speed is analyzed using 7uni-modal and six multi-modal functions. NL of MAOA converges more quickly than DE and BOA. NL error is smaller than BOA and DE. Rani et al. [13] have put out an NL topology for maritime search and rescue use. RSSI-based SSA is proposed, and the RSSI algorithm is used to compute the distance between the AN and TN. The value of the environment constant and path loss component is essential to the accuracy of node localization, and both values chosen for distance computation are not mentioned. SSA has 10% less NL error than BOA and PSO. The performance parameters are NL accuracy, localized nodes, NL errors, and computation time. Mohar et al. [40] have modified two Bat optimization algorithms (BOA) to localize the sensor nodes and overcome the drawbacks of the original BOA. Regarding NL error and computation time, NL based on BOA performs better than existing optimization techniques. The original BOA is modified, and the first variant of BOA is used to improve the global search strategy, and the second variant of BOA is used to improve the local search strategy. The performance of the proposed variants of BOA

is compared with the original BOA, PSO, GWO, BFOA, SSOA, and FOA. Muhammad Faheem et al. [43] proposed a scheme based on Industry 4.0 which is cost-effective and provides high-quality electricity. To ensure the quality of service (QoS) standards in the smart grid, efficient monitoring, and real-time management of unexpected changes in the power generating and distribution operations are crucial. Muhammad Faheem et al. [44] proposed a scheme based on Industry 4.0 and a huge data set derived from commercial Cyber-Physical Systems (CPS) with Internet of Things (IoT) support. A total of 450 sensors and 100 optical sensors are deployed to capture data from the real-time environment. The plan plays an essential role in improving system service performance to increase efficiency while improving customer experience. Muhammad Faheem et al. [45] presented an innovative method for data collection leveraging the smart grid's Internet of software-defined mobile sinks and WSNs. simulation is performed using EstiNet9.0. The developed system performs better than competing methods and accomplishes its specified objectives for events-driven applications in the SG. Packet delivery ratio, delay, throughput, and residual energy are the performance parameters. Similar kind of research work [46–52] is also caried out by the researchers to improve the performance of the proposed system.

Conclusion drawn from the related works Table 1 shows the summery of the related work. The accuracy of NL may significantly impact the performance of a WSN. The accuracy of the NL still needs to improve due to the path loss component and environment constant, which are not calibrated and whose values were chosen based on some assumptions. Authors have proposed numerous node-positing methods and meta-heuristic optimization schemes for accurate target node location. Appropriate path loss components and an environment constant are needed to enhance node localization. To further enhance NL, the meta-heuristic algorithm can be modified.

3 Background

In this section, the description of NL, SOA, chaotic maps, and levy flight problem is explained. As shown in Fig. 1, NL schemes can be broadly categorized into centralized and distrusted NL. Range-free and range-based are the distributed NL schemes [33–35, 39]. The range-based NL(RBNL) accuracy is low compared to range-free NL(RFNL) schemes. The position of the target nodes is estimated by RFNL using information about the network topology and connectivity to surrounding nodes. Range-based schemes perform the NL using either distance or angle. The combination of distance and angle can be used for NL, too. The Received Signal Strength Indicator (RSSI) [21–23] is the most widely used technique for estimating node position using node distance. Other location-aware strategies include Triangulation, Angle of Arrival (AoA), Time Difference of Arrival (TDoA), Time of Arrival (ToA), and Maximum Likelihood (ML) [32, 38].

3.1 Notation

Table 2 presents some abbreviation used in this work.

Table 1 Summery of the related work

Ref. & Year	Methodology/Algorithm	Research Gap	Performance Parameter
[12] 2021	GTOA	The optimization algorithm is mentioned only and the algorithm used for distance estimation between an anchor and target node is not mentioned	NL, NL Error
[25] 2021	RSSI	The RSSI method is used for distance estimation but the environment constant and path loss component value is essential to the accuracy of node localization and the constant value chosen lies between 2 to 5 The fixed value of path loss exponent 2.3 is chosen which may affect the NL accuracy	Average NL errors and computation time
[35] 2021	CS-GWO	All nodes are GPS enabled which consumes more power and cost-effective approach	NL, Mean NL error,
[36] 2021	QUATRE	Required some hardware and not cost-effective	NL accuracy, NL error
[37] 2020	MCB, PSO	There is a defect in MCB algorithms. Because all the candidate positions are generated randomly in an anchor box, no searching direction exists, which makes the time cost quite high and localization efficiency low. It may not be suitable for high mobility requirements	NL error, NL computation time
[2] 2019	SSA	The NL is primarily needed and method used for NL is not mentioned	NL accuracy, computation time and mean NL error
[3] 2019	DV-HOP, PSO	So many chaotic maps are available and a sine chaotic map is chosen for simulations and it is not mentioned why only a sine map is chosen for it	NL and NL Error
[18] 2022	MAOA, DV-Hop	This paper only discussed the application of node localization based on the two-dimensional plane in WSNs	Execution Time, Average NL error
[19] 2018	RSSI and Bat Algorithm	RSSI scheme is used for distance estimation between anchor node and target node but the value of environment constant and Path loss component value is not mentioned	Mean NL error, mean NL time
[3] 2017	GOA	The method used for NL is not mentioned	mean NL error, computation time, and NL
[4] 2017	BFOA	The trilateration method is used for distance computation but the value of environment constant and Path loss component value is not mentioned	Mean NL, mean NL error and mean computation time
[1] 2016	WOA	Covered only single objective function and multi-objective function is not dissuaded	Average cost function and standard deviation

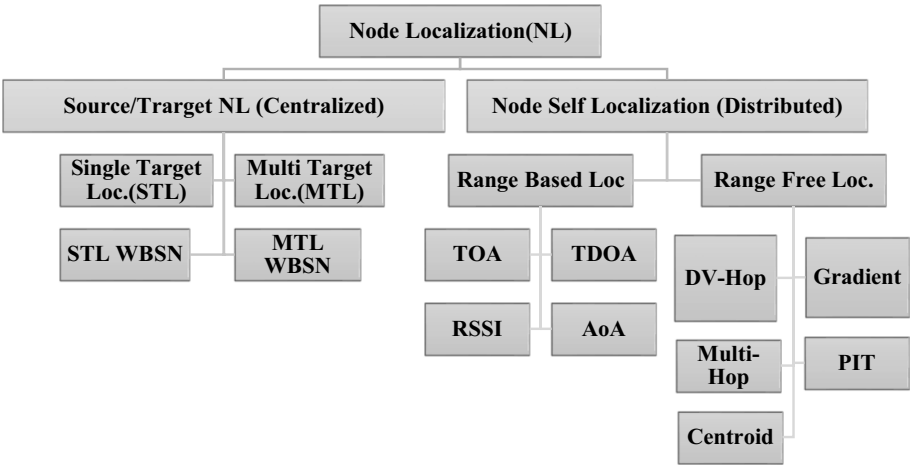


Fig. 1 NL Taxonomy

Table 2 Description of abbreviation

Abbreviation	Description
A	Variable representing motion behaviour
ANL_A	Average node localization accuracy
ANL_E	Average node localization error
ANL_T	Average node localization computation time
B	Random number
$C(t)$	New position of the search agent
$D(t)$	Distance covered by the Best-fit search agent
$d_{est(i)}$	Estimated distance between TN & AN
$d_{\mathrm{act}(i)}$	Actual distance between TN & AN
$itrMax$	Total Number of Iteration
f_c	Frequency constant
$M(t)$	Position of the search agent at time t
NL_E	Node localization error
$p_{best}(t)$	Position of the best search agent at time t
u	Correlation constant
v	Correlation constant
α	Step size
β	Chaotic index
r	Radius
rd	Random Number between 0 to 1
θ	Random angle between [0, 2π]
λ	Levy exponent

3.2 SOA

The SOA is a meta-heuristic algorithm shown in Algorithm 1.

Algorithm 1: Seagull Optimization Algorithm (SOA)
Input: Ps is the seagull population **Output**: Pbs is the best position for the search agent with the best fitness value using equ. (15) 1. Random initialization of *P(t), B*, and Maximum_ Iteration 2. Set the value of the *fc*=2, u=1, and *v*=1, *t*=0. 3. **While** *(t*<Maximum_Iteration) 4. { Compute Fitness value($P(t)$) 5. Generate the random number(rd)=(0, 1) 6. Generate the random number(θ)= (0, 2π) 7. Compute$r = u * e^{\theta v}$ 8. Estimate the distance *(D(t))* 9. Estimate the new position (p_{best}) 10. Update the position of the optimal seagull and fitness value *11. t=t+1;* 12. } End while 13.**Output**: Best position (p_{best}) of the seagull and fitness value

The mathematical model of SOA [29] mainly depends on exploration and exploitation. The exploration phase is accomplished with three conditions. Collision avoidance, movements toward the best neighbor's direction, and being close to the best search agent. To avoid collisions with other seagulls, variable *A* is employed to estimate the new position of the search agent(seagull) as follows:

$$C(t) = A * P(t); \tag{1}$$

where *C(t)* is the new position of the search agent and *P(t)* is the current position of the other search agent that does not collide with another search agent at iteration *t*.

A represents the motion behavior of the search agent in each search space. *A* is a variable that represents the motion behavior of the search agent in each search space. The behavior of variable *A* is controlled using the frequency constant*(fc)*. The value of *fc* is 2.

$$A = fc - t * \left(\frac{fc}{itrMax}\right) \tag{2}$$

where tis the iteration from 0 to itrMax, and the value of *A* decreases from *fc* to 0.

Then search agent moves in the direction of the best solution.

$$M(t) = B - P_{best}(t) - P(t) \tag{3}$$

where *M(t)* is the position of the search agent *P(t)* towards the best search agent $p_{best}(t)$. *B* is a random number that balances the local and global search agents.

$$B = 2 * A^2 - rd \tag{4}$$

where the random number (rd) lies between 0 to 1.

Now search agent moves in the direction of the best search agent.

$$D(t) = M(t) + C(t) \tag{5}$$

where *D(t)* represents the best-fit search agent.

The search agents can change the speed and angle for attacking the prey. While attacking the prey, the search agent moves in a spiral shape. The motion behavior of the agent 3-D planes is as follows:

$$(x, y, z) = \left(r * \cos\theta,\ r* \sin\theta,\ r * \theta\ \ wherer = u * e^{\theta v} \right) \tag{6}$$

where the radius spiral is r and θ is a random angle between [0, 2π].

The correlation constants are u and v of the spiral shape, and e is the base of the natural logarithm. The attack position of the search agent is constantly updated.

$$P(t) = D(t) * x * y * z + p_{best}(t) \tag{7}$$

where P(t) saves the best solution and updates the position of other search agents.

3.3 Chaotic maps

The meaning of chaos [6, 10, 14–17] is defined by properties of deterministic, nonlinear, semi-stochastic, and sensitive dependence. For best results, choosing the initial values carefully is crucial because chaotic maps are very sensitive to changes in the initial conditions. Chaos is the randomness of straightforward deterministic dynamic systems in mathematics, and chaotic systems can be thought of as sources of randomness. Chaos employs chaotic variables rather than random variables. Ergodicity and non-repetition are other valuable properties that enhance the search process. Chaotic can avoid reaching local maxima. Although the meta-heuristic optimization algorithm is robust, efficient, flexible, and very popular in the research community, it has a slow convergence speed. To improve the performance of the optimization scheme, the logistic chaotic map shown in Fig. 2 is found to be more suitable out of the 10 chaotic maps shown in Table 3.

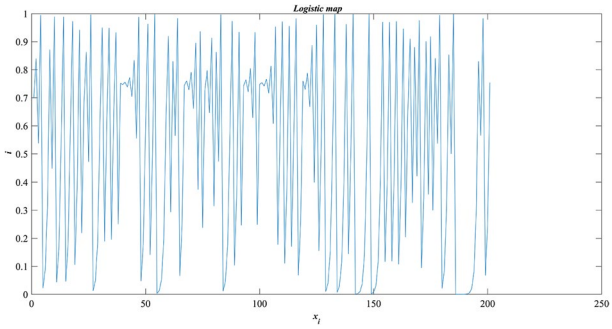


Fig. 2 Logistic map of the chaotic function

Table 3 List of chaotic maps

Sr.NO	Name	Chaotic map	Constant	Range
1	Tent map	$x_{i+1} = \begin{cases} \mu x_i, for x_{i<0.5} \\ \mu(1-x)_i, for x_{i\geq 0.5} \end{cases}$	$\mu = 2$	(0,1)
2	Logistic map	$x_{i+1} = \frac{\mu}{4} x_i (1 - x_i)$	$0 \leq \mu \leq 4$	(0,1)
3	Sine map	$x_{i+1} = \frac{\mu}{4} sin\pi x_i$	$0 < \mu \leq 4$	(0,1)
4	Kent map	$x_{i+1} = \begin{cases} \frac{x_i}{m}, & 0 < x_i < m \\ 1 - x_i, & m < x_i < 1 \end{cases}$	$0<m<1$	(0,1)
5	Circle map	$x_{i+1} = x_i + \mu - \frac{\tau}{2\pi} \sin 2\pi x_i mod 1$	$\mu = 0.5, \tau = 0.2$	(0,1)
6	Cubic map	$x_{i+1} = c x_i (1 - x_i^2)$	C=2.59	(0,1)
7	Singer map	$x_{i+1} = \mu(7.86x_i - 23.31x_i^2 + 28.75x_i^3 - 13.3x_i^4)$	$0.9 \leq \mu \leq 1.08$	(0,1)
8	Gauss map	$x_{i+1} = \begin{cases} 1, if x_i = 0 \\ \frac{1}{x_i mod 1} \quad otherwise \end{cases}$	$\mu = 6$	(0,1)
9	Square map	$x_{i+1} = (1 - \mu x_i^2)$	$\mu = 2$	
10	Sinusoidal map	$x_{i+1} = \alpha x_i^2 sin\pi x_i$	$\alpha = 2.3$	(0,1)

Randomness is the basic property of the meta-hubristic optimization algorithms, which is modified using a chaotic map to improve the performance of the algorithm. A total 200 iterations are performed for the generation of the random number between 0 to 1.

3.4 Levy flight

Levy flight is just a random walk with step lengths that adhere to the levy distribution, a probability distribution with a heavy tail depicted in Fig. 3. Levy distribution is a series of small steps. When a walk is said to have a dimension more than once, the directions in which the steps are taken are isotropic. Lévy-flight can improve population diversity to speed up the convergence and reduce the local optimum. Better exploitation and exploration balance is made possible with the help of this strategy. A seagull follows the levy flight distribution. The levy flight can suddenly take a more extended step size.

$$p_i^{t+1} = p_i^t a \otimes levy(\lambda) \tag{8}$$

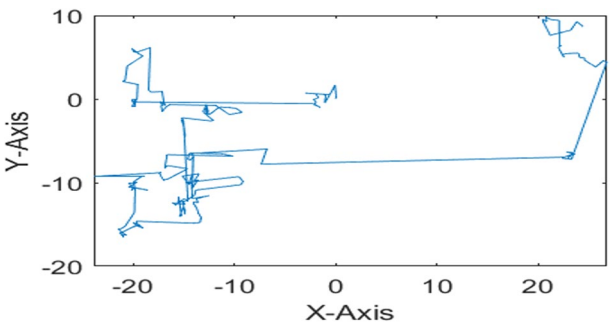


Fig. 3 Randomness of levy flight

where p_i^{t+1} is the search agent's new position, p_i^t is the search agent's current position, α is the step size and $levy(\lambda)$ is the levy exponent. The step size σ calculates how far a search agent can go for a fixed number of iterations. The step size affects the search agent's rate of convergence. The new position of the search agent would be very close to the present position if the step size is small, and it would be far from the current position if the step size is large. The step size of the levy flight is as follows.

$$\alpha = \frac{\sigma_u * u}{|v|^{1/\beta}} \tag{9}$$

where the variables u and v are stochastic. The values of the variables u and v, which follow the normal distribution, are as follows.

$$\mathrm{u} = rand.\sigma_u, 0 < u < \propto_u^2 \tag{10}$$

$$v = rand,\ 0 < v < \propto_v^2 \tag{11}$$

The standard deviation(σ_u) is as follows.

$$\sigma_u = \left(\frac{\gamma(1+\beta).\sin\left(\pi.\frac{\beta}{2}\right)}{\gamma\left(\frac{1+\beta}{2}\right).\beta.2^{\frac{\beta-1}{2}}} \right)^{\frac{1}{\beta}} \tag{12}$$

The value of $\sigma_v = 1$ and $\beta = 3/2$ are considered.

4 The proposed enhancements over the seagull algorithm

The standard SOA has been modified in this section based on two mathematical techniques, levy-flight, and chaos theory. The SOA has the drawbacks of a slow convergence rate, entrapment in local minima, and skipping real solutions throughout the optimization process. Two methods have been used to address the afore mentioned problems. The second technique has used Levy flight distribution to address the typical SOA's diversity problem. The local optima problem can be solved with the help of the levy distribution's infinite variance and changeable step size, which broadens the search process's scope and diversity.

4.1 Proposed Chaotic map-based seagull optimization algorithm (C-SOA)

The working of C-SOA is illustrated in algorithm 2. The target and anchor nodes are deployed in the environment using a chaotic pseudo-random number generation technique. The fitness value of each search agent(seagull) is estimated using the fitness function. Estimate the distance between the search agent and TN. Update the position of the seagull till the target node is not localized. Figure 4 shows the pseudo-code of C-SOA.

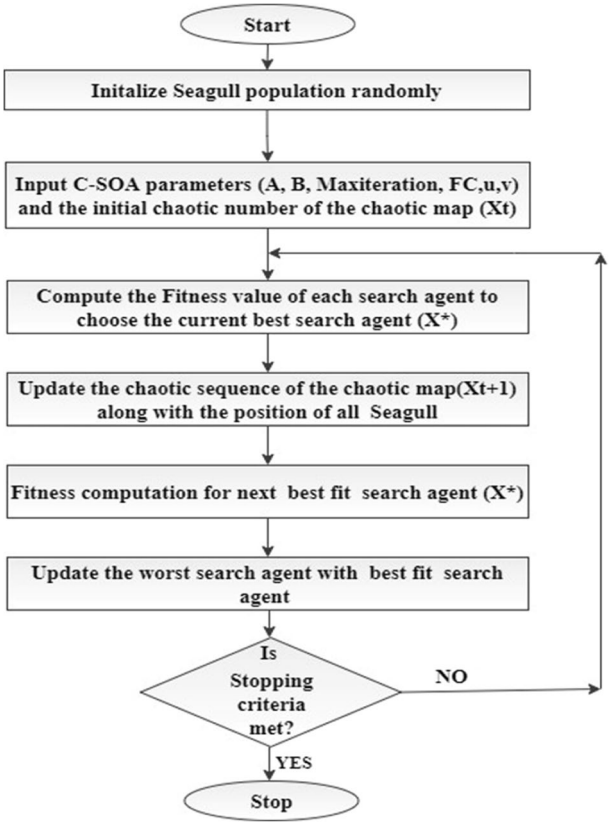


Fig. 4 Flowchart for C-SOA-based NL in WSNs

Algorithm-2: Chaotic map-based Seagull Optimization Algorithm (C-SOA)
Input: Ps is the seagull population using chaotic pseudo-random number (c_r)
Output: p_{best} is the best position for the search agent (seagull)
1. Random initialization of *P(t), B*, and Maximum_ iteration
2. Set the value of the *fc=2, u=1,* and *v=1, t=0.*
3. **While** (t<Maximum_Iteration)
4. { Compute the Fitness value (P_t) of each seagull to choose the current best
5. Generate the random number(rd)=(0, 1)
6. Generate the random number(θ)= (0, 2π)
7. Compute$r = u * e^{\theta v}$
8. Estimate the distance (*D(t)*)
9. Estimate the new position (p_{best})
10. Update the position of the optimal seagull along with c_r and fitness value
11. *t=t+1;*
12. } **End while**
13. Output: Best position (p_{best}) of the seagull and fitness value

4.2 Proposed Levy flight for the SOA (LF-SOA)

In Levy flight, there is a broad distribution of a single step of lengths, but each step is in the random direction, where the most extended single step of another random walk can dominate the random walk after many steps. Levy flight is a mathematical algorithm that mimics seagull hunting behavior. Lévy-flight distribution has been embedded with the seagull optimization algorithm. The most effective method to arbitrarily search for a target in an environment is represented by Lévy-flights. The proposed algorithm name is LF-SOA, shown in Fig. 5.

The Sensor nodes are deployed in the environment randomly. The fitness value of each search agent is estimated. Estimate the distance between the search agent and TN, and change the position of the search agent till the nodes are not localized. Algorithm 3 shows the pseudo-code of the LF-SOA.

Algorithm 3: Levy Flight-based Seagull Optimization Algorithm (LF-SOA)
Input: Ps is the seagull population deployed using levy flight. Output: Pbs is the best position for the search agent (seagull) 1. Random initialization of *P(t), B*, and Maximum_ Iteration 2. Set the value of the *fc=2, u=1,* and *v=1, t=0.* 3. **While** (t<Maximum_Iteration) 4. { Compute the Fitness value(P_t) of each seagull to choose the current best 5. Generate the random number(rd)=(*0, 1*) 6. Generate the random number(θ)= (*0, 2π*) 7. Compute$r = u * e^{\theta v}$ 8. Estimate the distance (*D(t)*) 9. Estimate the new position (p_{best}) 10. Update the position of the optimal seagull along with c_r and fitness value *11.* *t=t+1;* 12. } **End while** 13. Output: Best position (p_{best}) of the seagull and fitness value

5 Methodology for node localization

The architecture of the NL is shown in Fig. 6. The main objective is to find the location of the unknown nodes. There are two steps in RBNL: the distance estimation between AN and TN, followed by the location estimation. Each TN estimates its distance from the ANs using the RSSI value. Then, the location of the target nodes is estimated using the distance measured between the AN and the TN. It can be done using either a

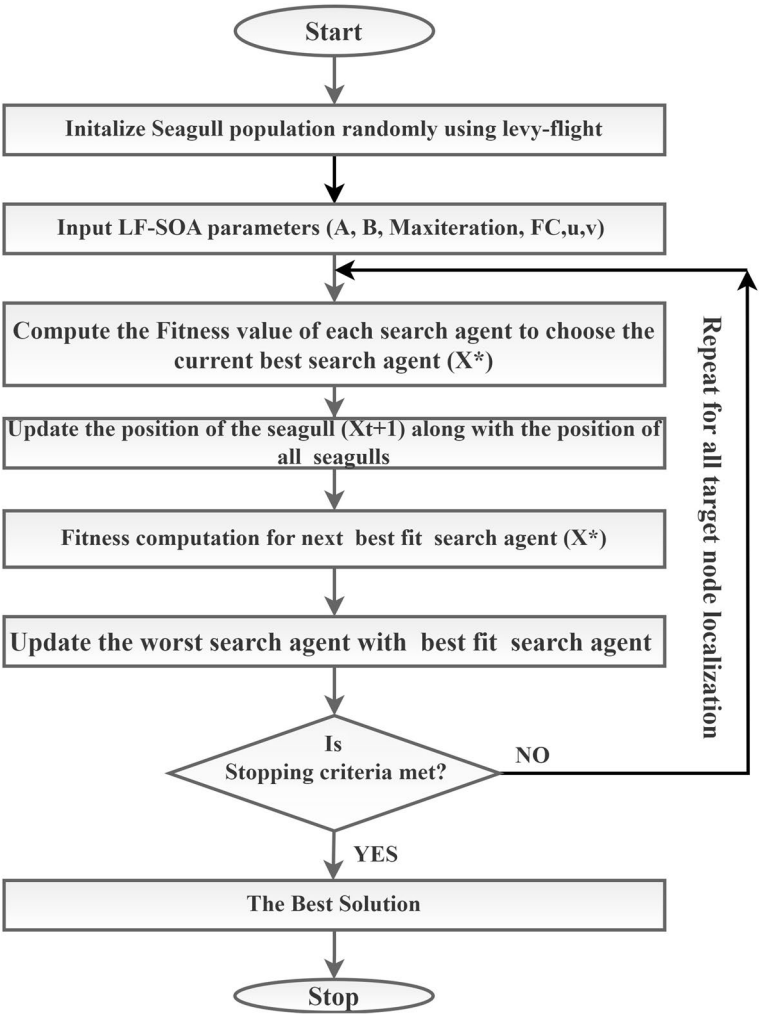


Fig. 5 Flowchart for LF-SOA-based NL in WSNs

geometric approach or an optimization algorithm. Several metaheuristic methodologies have been introduced to address the NL problems. The following process is being used for the NL.

5.1. Deploy AN and TN randomly in the sensor field.
5.2. Estimate the distance between TN (xt, yt) and AN (x, y) using Euclidian distance measurement methods as follows.

$$d_{est} = \sqrt{(x - x_t)^2 + (y - y_t)^2} \tag{13}$$

The distance estimated between the nodes by RSSI contains some errors. The estimated distance (d_{est}) has been affected by the noise(ni), which is in the range of $d_{est} \pm d_{est}(\frac{n_0}{100})$, where n_0 is the noise in the measured distance.

5.3. If the TN is placed within the communication range of three or more anchor nodes, the node is said to be localizable. The position of the TN is estimated using the trilateration method.
5.4. The SOA, C-SOA, and LF-SOA are used to find the optimal position of the TN. Seagull is the search agent in the NL scheme. All the search agents are initialized with centroids of the ANs available within the communication range. The centroid (x_{ck}, y_{ck}) of the ANs that exist within the communication range of the localizable node is as follows.

$$(x_{ck}, y_{ck}) = \left(\frac{1}{n}\sum\nolimits_i^n x_i, \frac{1}{n}\sum\nolimits_i^n y_i\right) \tag{14}$$

5.5. The actual positions of the sensor nodes in simulations are known, so it is possible to determine the actual distances ($d_{act(i)}$) between nodes AN and TN. The objective function for the NL is the mean square distance between AN and TN, which is minimized using the SOA, C-SOA, and LF-SOA. The mean square distance between AN and TN is as follows.

$$f(x, y) = \frac{1}{n}\left(\sum\nolimits_{i=1}^n d_{est(i)} - d_{act(i)}\right)^2, \mathrm{n} \geq 3 \tag{15}$$

5.6. The SOA, C-SOA, and LF-SOA are used to minimize the objective function, which reduces the NL error, and the optimal value of TN is found.
5.7. Repeat the steps from 5.2 to 5.6 till all TNs are not localized.
5.8. NL error (NL_E) is estimated after all target nodes are localized. NL_E between all ANs (xi, yi) and TN (Xi, Yi) are as follows.

$$NL_E = \frac{1}{n}\sum\nolimits_{i=1}^n \sqrt{(x_i - X_i)^2 + (y_i - Y_i)^2} \tag{16}$$

The performance of the SOA, C-SOA, and LF-SOA is compared using average NL error (ANL_E), NL computation time(NL_T), and the number of TNs localized.

NL optimization NL optimization is shown in Fig. 7. The AN & TN are deployed in an area of 200×200 m. The distance between the AN and TN is estimated using the RSSI method, and the position of the TN is computed using three ANs. The fitness value of each search agent is computed, and the mean square error is estimated. Finally, the TNs are localized.

6 Performance evaluation and result discussion

The simulation parameters required to simulate the proposed work is presented in Table 4. Performance metrics of SOA, C-SOA, and LF-SOA is presented in Table 5. The average value of the NL_E(m), NL_T(s), and NL_A is estimated and presented in Table 6.

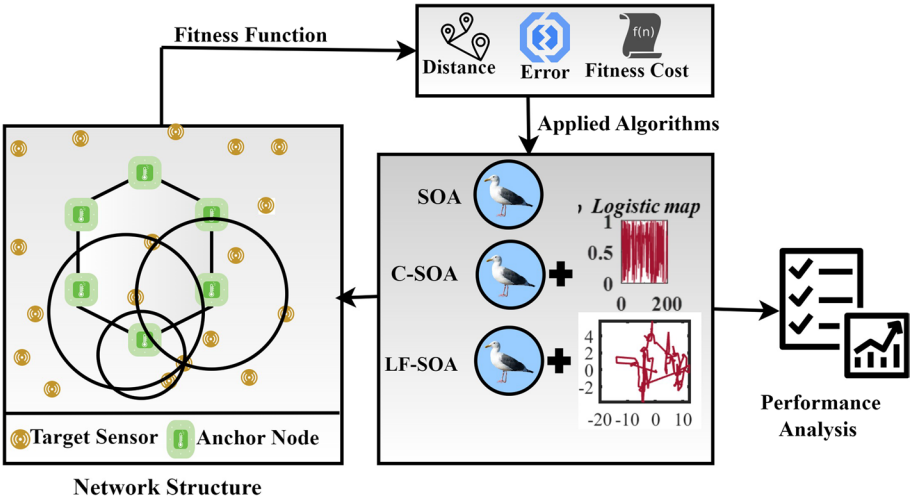


Fig. 6 Architecture of NL in WSNs

6.1 Simulation environment

The simulation experiment is performed on an Intel(R) Core (TM) i7-1165G7 processor running at 2.80 GHz, 2803 MHz, with 4 Cores and 8 Logical Processor (s), and MATLAB2021a simulation environment. In all simulations, the coordinates of the ANs and TNs are randomly configured during the construction of the deployment area. All the sensor nodes deployed in the field are static throughout the simulations. The main task is to localize unknown sensor nodes in a 2-D environment. The environmental area for the simulation is 200×200 m. A total of six cases are taken in which the size of ANs varies from 15 to 90, and TNs varies from 25 to 150.

Node localization If there are not enough ANs, it is difficult to locate the position of the TNs. With an increase in anchor node density, NL accuracy increases. The NL accuracy depends on the number of anchor nodes used for NL in LF-SOA, C-SOA, and SOA, as shown in Fig. 8.The accuracy of NL likewise increases with the number of iterations.

Average node localization error Figure 9 shows the comparative results among the SOA, C-SOA, and LF-SOA, and the average NL error is estimated. The simulation result shows that the average node localization error of LF-SOA is the least. When the size of AN and TN are 15 and 25, respectively, the average node localization error is 0.41. When the size of AN and TN are increased to 90 and 150, respectively, the average node localization error is reduced to 0.271, which is significantly less. Hence, the NL accuracy increases if the AN density is increased. The NL error decreases as the number of iterations increases. LF-SOA exhibits a greater drop in error than the SOA and C-SOA.

Computation time The average node localization time is another performance metric shown in Fig. 10, and LF-SOA has faster node localization than the other two schemes. In contrast to SOA and C-SOA, the computation time for LF-SOA is extremely short.

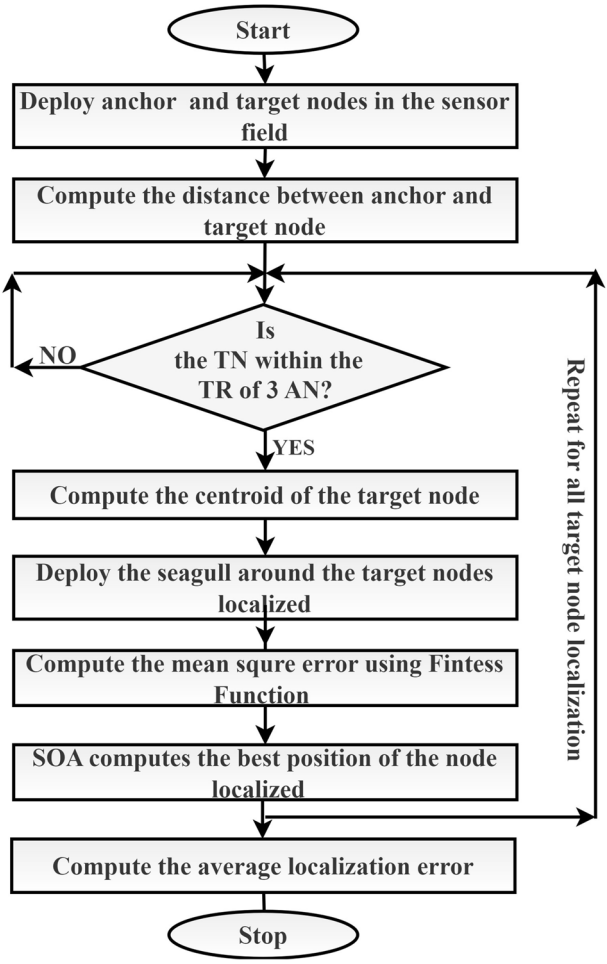


Fig. 7 Flowchart for NL optimization

Table 4 Simulation parameters for SOA, C-SOA, and LF-SOA

Algorithm	Parameter	Value	Unit
	Target nodes	25,50,75,100,125,150	
	Anchor nodes	15, 30,45,60,75,90	
	Node transmission range (R)	30	m
	Deployment area	200×200	m^2
	Total Iteration	20,40,60,80,100	
SOA	Frequency-controlled parameter (f_c)	2	
C-SOA	Chaotic logistic map constant (μ)	4	
	Chaotic Index (β)	1.5	
LF-SOA	Correlation constant (u)	1	
	Correlation constant (v)	1	

Table 5 Performance metrics of SOA, C-SOA, and LF-SOA

TN	AN	Iterations	SOA			C-SOA			LF-SOA		
			NL_E(m)	NL_T(s)	NL_A	NL_E(m)	NL_T(s)	NL_A	NL_E(m)	NL_T(s)	NL_A
25	15	20	0.718	0.38	13	0.628	0.32	15	0.421	0.35	18
		40	0.714	0.38	12	0.611	0.35	16	0.418	0.36	19
		60	0.706	0.40	14	0.602	0.39	16	0.409	0.30	20
		80	0.692	0.41	15	0.593	0.40	18	0.401	0.38	22
		100	0.652	0.40	16	0.583	0.40	19	0.401	0.38	23
50	30	20	0.662	0.47	40	0.543	0.45	40	0.398	0.41	44
		40	0.655	0.49	41	0.535	0.45	42	0.396	0.42	46
		60	0.649	0.49	43	0.527	0.42	45	0.396	0.42	47
		80	0.641	0.50	44	0.519	0.42	45	0.391	0.45	48
		100	0.641	0.51	44	0.510	0.40	46	0.386	0.45	48
75	45	20	0.589	0.66	58	0.502	0.57	66	0.387	0.48	70
		40	0.581	0.68	59	0.493	0.57	67	0.382	0.48	71
		60	0.577	0.68	59	0.489	0.51	67	0.382	0.51	71
		80	0.567	0.69	61	0.474	0.49	69	0.374	0.52	73
		100	0.566	0.70	62	0.474	0.48	72	0.369	0.52	74
100	60	20	0.538	0.74	85	0.449	0.56	93	0.361	0.56	94
		40	0.531	0.76	87	0.441	0.61	94	0.361	0.57	96
		60	0.527	0.77	87	0.433	0.61	94	0.350	0.57	97
		80	0.521	0.79	89	0.426	0.65	95	0.342	0.61	97
		100	0.511	0.79	92	0.426	0.66	96	0.337	0.62	99
125	75	20	0.501	0.83	110	0.417	0.68	118	0.320	0.72	121
		40	0.491	0.85	110	0.409	0.77	120	0.311	0.73	121
		60	0.487	0.86	113	0.399	0.79	121	0.309	0.73	123
		80	0.481	0.89	112	0.377	0.79	121	0.296	0.76	124
		100	0.473	0.93	115	0.377	0.81	122	0.296	0.77	124
150	90	20	0.451	1.19	138	0.368	0.93	143	0.289	0.89	146
		40	0.447	1.27	139	0.351	0.93	143	0.289	0.89	146
		60	0.447	1.27	140	0.347	0.96	145	0.271	0.91	147
		80	0.427	1.30	141	0.336	0.98	145	0.264	0.93	148
		100	0.419	1.33	142	0.336	0.99	147	0.259	0.93	149

Table 6 Result summery of SOA, C-SOA, LF-SOA

TN	AN	SOA			C-SOA			LF-SOA		
		ANL_E(m)	ANL_T(s)	ANL_A	ANL_E(m)	ANL_T(s)	ANL_A	ANL_E(m)	ANL_T(s)	ANL_A
25	15	0.696	0.39	16	0.603	0.37	17	0.410	0.35	21
50	30	0.649	0.49	43	0.526	0.42	44	0.393	0.43	47
75	45	0.567	0.68	64	0.486	0.52	69	0.378	0.55	72
100	60	0.525	0.77	88	0.435	0.61	95	0.350	0.67	97
125	75	0.486	0.87	116	0.395	0.76	121	0.306	0.81	123
150	90	0.438	1.27	143	0.347	0.95	145	0.274	0.91	148

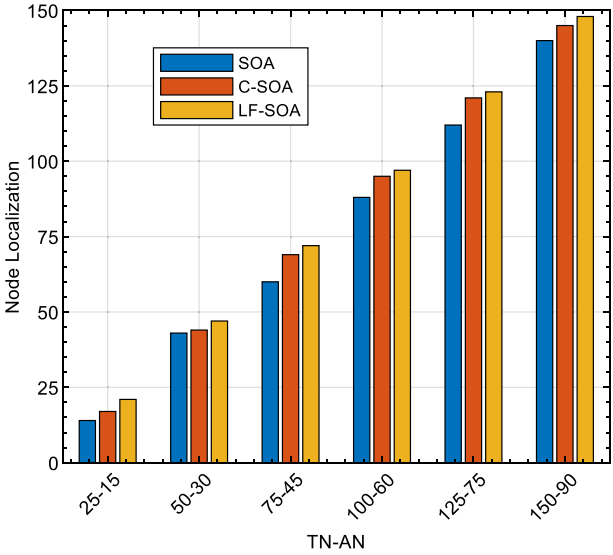


Fig. 8 Variation of Average Node Localization with TN-AN at different optimization algorithm

7 Conclusion

The modified SOA is proposed for NL in the present article. The primary goal of this optimization is to reduce the localization error, minimize the computation time, and enhance the accuracy of NL. The conventional SOA is modified using the fitness function and simulation is caried out on modified SOA. Modified SOA is further modified using the chaotic map mentioned above is known as C-SOA and simulation is carried out with the proposed fitness function. Modified SOA used Levy-flight for randomization and simulation is carried out again with the proposed fitness function. The comparative analysis among SOA,

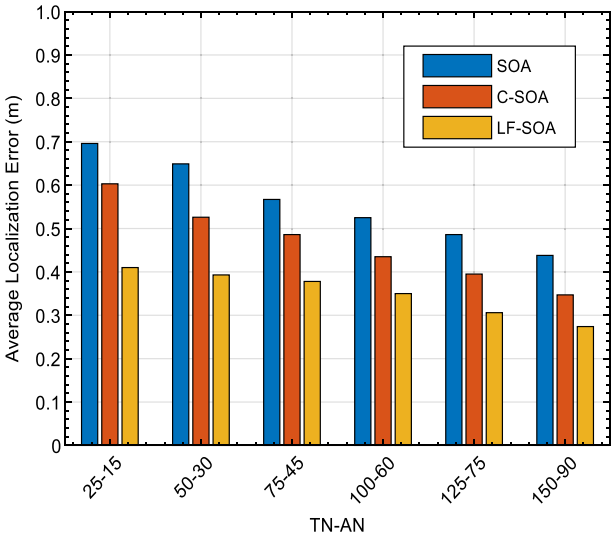


Fig. 9 Variation of Average Node Localization Error (ANL_E) with TN-AN at different optimization algorithm

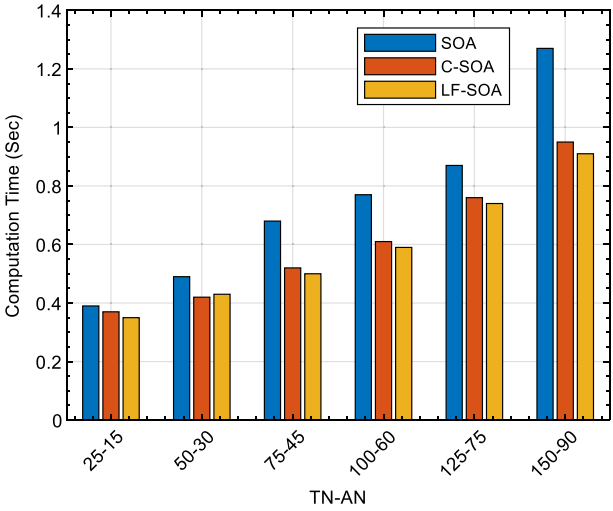


Fig. 10 Variation of computation time with TN-AN at different optimization algorithm

C-SOA, and LF-SOA is performed using the proposed simulation condition. The outcomes of the simulations demonstrate how well the suggested method refines NL. Regarding error and computation time, the C-SOA and LF-SOA do far better than the SOA, and the performance of the LF-SOA is better than the C-SOA. The proposed C-SOA and LF-SOA are applied to a range-based NL scheme.

7.1 Future work

The future work will investigate the range-free NL scheme to minimize NL error, minimization of computation time, and optimization of NL accuracy. Further, SOA can be hybridized with another meta-heuristic algorithm to optimize the above performance parameters.

Funding No funding was received to assist with the preparation of this paper.

Data availability All relevant data and materials related to the paper have been included and properly cited.

Declarations

Conflict of interest Regarding publishing this paper, the authors affirm that there are no conflicts of interest.

References

1. Mirjalili S, Lewis A (2016) The whale optimization algorithm. Adv Eng Softw 95:51–67
2. Kanosh HM, Houssein EH, Selim MM (2019) Salp swarm algorithm for node localization in wireless sensor networks. J Comput Netw Commun 2019
3. Rajakumar R, Amudhavel J, Dhavachelvan P, Vengattaraman T (2017) GWO-LPWSN: Grey wolf optimization algorithm for node localization problem in wireless sensor networks. J Comput Netw Commun 2017
4. Arora S, Singh S (2017) Node localization in wireless sensor networks using butterfly optimization algorithm. Arab J Sci Eng 42(8):3325–3335

5. Sun Z, Tao Li, Wang X, Zhou Z (2015) Localization algorithm in wireless sensor networks based on multi-objective particle swarm optimization. Int J Distrib Sens Netw 11(8):716291
6. Feng J, Zhang J, Zhu X, Lian W (2017) A novel chaos optimization algorithm. Multimed Tools Appl 76(16):17405–17436
7. Arora S, Singh S (2019) Butterfly optimization algorithm: a novel approach for global optimization. Soft Comput 23(3):715–734
8. Tuba E, Tuba M, Beko M (2018) Two stage wireless sensor node localization using firefly algorithm. Smart trends in systems, security and sustainability. Springer, Singapore, pp 113–120
9. Lee M, Yoe H (2019) WiBiA: Wireless sensor networks based on biomimicry algorithms. Int J Comput Intell Syst 12(2):1212–1220
10. Demidova LA, Gorchakov AV (2020) A study of chaotic maps producing symmetric distributions in the fish school search optimization algorithm with exponential step decay. Symmetry 12(5):784
11. Li Tu, Yan W, Ping Li, Fang P (2019) A WSN positioning algorithm based on 3D discrete chaotic mapping. EURASIP J Wirel Commun Netw 2019(1):1–13
12 Sekhar P, Laxmi Lydia E, Elhoseny M, Al-Akaidi M, Selim MM, Shankar K (2021) An effective metaheuristic-based node localization technique for wireless sensor networks enabled indoor communication. Phys Commun 48:101411
13. Shalli R, Babbar H, Kaur P, Alshehri MD, Shah SHA (2022) An optimized approach of dynamic target nodes in wireless sensor network using bio inspired algorithms for maritime rescue. IEEE Trans Intel Transp Syst
14. Arora S, Anand P (2019) Chaotic grasshopper optimization algorithm for global optimization. Neural Comput Appl 31(8):4385–4405
15. Kaur G, Arora S (2018) Chaotic whale optimization algorithm. J Comput Des Eng 5(3):275–284
16. Tavazoei MS, Haeri M (2007) Comparison of different one-dimensional maps as chaotic search pattern in chaos optimization algorithms. Appl Math Comput 187(2):1076–1085
17 Alwan NAS, Hussain ZM (2019) Compressive sensing with chaotic sequences: an application to localization in wireless sensor networks. Wirel Pers Commun 105(3):941–950
18 Cheng M, Qin T, Yang J (2022) Node localization algorithm based on modified archimedes optimization algorithm in wireless sensor networks. J Sensors 2022:1
19. Mihoubi M, Rahmoun A, Lorenz P (2017) "Metaheuristic RSSI based for node localization in distributed wireless sensor network." In 2017 Global Information Infrastructure and Networking Symposium (GIIS). IEEE, pp 64–70
20. Kulkarni RV, Venayagamoorthy GK (2010) Particle swarm optimization in wireless-sensor networks: A brief survey. IEEE Trans Syst, Man, Cybern, Part C (Appl Rev) 41(2):262–267
21. Huang Y, Zheng J, Xiao Y, Peng M (2015) Robust localization algorithm based on the RSSI ranging scope. Int J Distrib Sens Netw 11(2):587318
22. Zhang Y, Xing S, Zhu Y, Yan F, Shen L (2017) RSS-based localization in WSNs using Gaussian mixture model via semidefinite relaxation. IEEE Commun Lett 21(6):1329–1332
23. Ismail MIM, Dzyauddin RA, Samsul S, Azmi NA, Yamada Y, Yakub MFM, Binti NA, Salleh A (2019) "An RSSI-based wireless sensor node localisation using trilateration and multilateration methods for outdoor environment." arXiv preprint arXiv:1912.07801
24. Xu Y, Zhou J, Zhang P (2014) RSS-based source localization when path-loss model parameters are unknown. IEEE Commun Lett 18(6):1055–1058
25. Chuku N, Nasipuri A (2021) RSSI-Based localization schemes for wireless sensor networks using outlier detection. J Sens Actuator Netw 10(1):10
26. Zafari F, Gkelias A, Leung KK (2019) A survey of indoor localization systems and technologies. IEEE Commun Surv Tutorials 21(3):2568–2599
27. Al Shayokh M, Shin SY (2017) Bio inspired distributed WSN localization based on chicken swarm optimization. Wirel Pers Commun 97(4):5691–5706
28. Phoemphon S, So-In C, Leelathakul N (2021) Improved distance estimation with node selection localization and particle swarm optimization for obstacle-aware wireless sensor networks. Expert Syst Appl 175:114773
29. Dhiman G, Kumar V (2019) Seagull optimization algorithm: Theory and its applications for large-scale industrial engineering problems. Knowl-Based Syst 165:169–196
30. Mihoubi M, Rahmoun A, Lorenz P, Lasla N (2018) An effective Bat algorithm for node localization in distributed wireless sensor network. Secur Priv 1(1):e7
31. Strumberger I, Minovic M, Tuba M, Bacanin N (2019) Performance of elephant herding optimization and tree growth algorithm adapted for node localization in wireless sensor networks. Sensors 19(11):2515

32. Wang W, Liu X, Li M, Wang Z, Wang C (2019) Optimizing node localization in wireless sensor networks based on received signal strength indicator. IEEE Access 7:73880–73889
33 Lalama Z, Boulfekhar S, Semechedine F (2022) Localization optimization in WSNS using meta-heuristics optimization algorithms: a survey. Wirel Pers Commun 122(2):1197–1220
34. Sharma N, Gupta V (2020) Meta-heuristic-based optimization of WSNs localization problem-a survey. Procedia Comput Sci 173:36–45
35. Arafat MY, Moh S (2021) Bio-inspired approaches for energy-efficient localization and clustering in UAV networks for monitoring wildfires in remote areas. IEEE Access 9:18649–18669
36 Song L, Jiang X, Wang L, Xiaochun H (2021) Monte Carlo node localization based on improved QUARTE optimization. J Sensors 2021:1
37 Wu H, Liu J, Dong Z, Liu Y (2020) A hybrid mobile node localization algorithm based on adaptive MCB-PSO approach in wireless sensor networks. Wirel Commun Mob Comput 2020:1
38. Sankalap A, Kaur R (2017) Nature inspired range based wireless sensor node localization algorithms
39. Cheng L, Chengdong Wu, Zhang Y, Hao Wu, Li M, Maple C (2012) A survey of localization in wireless sensor network. Int J Distrib Sens Netw 8(12):962523
40. Mohar SS, Goyal S, Kaur R (2022) Localization of sensor nodes in wireless sensor networks using bat optimization algorithm with enhanced exploration and exploitation characteristics. J Supercomput 78(9):11975–12023
41. Derr K, Manic M (2015) Wireless sensor networks—Node localization for various industry problems. IEEE Trans Industr Inf 11(3):752–762
42. Dohare I, Singh K (2020) Green communication in sensor-enabled IoT: integrated physics inspired meta-heuristic optimization-based approach. Wireless Netw 26(5):3331–3348
43. Faheem M, Fizza G, Ashraf MW, Butt RA, Ngadi MA, Gungor VC (2021) Big Data acquired by Internet of Things-enabled industrial multichannel wireless sensors networks for active monitoring and control in the smart grid Industry 4.0. Data Brief 35:106854
44. Faheem M, Butt RA (2022) Big datasets of optical-wireless cyber-physical systems for optimizing manufacturing services in the internet of things-enabled industry 4.0. Data Brief 42:108026
45. Faheem M, Umar M, Butt RA, Raza B, Ngadi MA, Gungor VC (2019) "Software defined communication framework for smart grid to meet energy demands in smart cities." In 2019 7th International Istanbul Smart Grids and Cities Congress and Fair (ICSG). IEEE, pp 51–55
46. Jiang T, Yang X, Yang Y, Chen Xi, Bi M, Chen J (2022) Wavelet method optimised by ant colony algorithm used for extracting stable and unstable signals in intelligent substations. CAAI Trans Intell Technol 7(2):292–300
47. Yuli D, Zeng Z, Kritshekhar J, Huang D (2021) Problem-based cybersecurity lab with knowledge graph as guidance. J Artif Intel Tech
48. Khamkar R, Das P, Namasudra S (2023) SCEOMOO: a novel subspace clustering approach using evolutionary algorithm, off-spring generation and multi-objective optimization. Appl Soft Comput 139:110185
49. Gao X, Cao W, Yang Q, Wang H, Wang X, Jin G, Zhang J (2022) Parameter optimization of control system design for uncertain wireless power transfer systems using modified genetic algorithm. CAAI Trans Intell Technol 7(4):582–593
50. Namasudra S, Nath S, Majumder A (2014) "Profile based access control model in cloud computing environment." In 2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE). IEEE, pp 1–5
51. Wang M, Hualing Y, Feng J, Ling L, Min G (2022) Review on offloading of vehicle edge computing. J Artific Intell Technol
52. Malviya S, Kumar P, Namasudra S, Tiwary US. (2022) Experience replay-based deep reinforcement learning for dialogue management optimisation. Trans Asian and Low-Resource Lang Inf Process

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.

Electronic Research Archive

ERA, 32 (2): 1268−1307.
DOI: 10.3934/era.2024060
Received: 27 February 2023
Revised: 28 May 2023
Accepted: 06 July 2023
Published: 31 January 2024

http://www.aimspress.com/journal/ERA

Research article

Securing cloud-enabled smart cities by detecting intrusion using spark-based stacking ensemble of machine learning algorithms

Mohd. Rehan Ghazi* and N. S. Raghava

Department of Electronics and Communication Engineering, Delhi Technological University, Delhi 110042, India

* **Correspondence:** Email: er.rehan.aras@gmail.com.

Abstract: With the use of cloud computing, which provides the infrastructure necessary for the efficient delivery of smart city services to every citizen over the internet, intelligent systems may be readily integrated into smart cities and communicate with one another. Any smart system at home, in a car, or in the workplace can be remotely controlled and directed by the individual at any time. Continuous cloud service availability is becoming a critical subscriber requirement within smart cities. However, these cost-cutting measures and service improvements will make smart city cloud networks more vulnerable and at risk. The primary function of Intrusion Detection Systems (IDS) has gotten increasingly challenging due to the enormous proliferation of data created in cloud networks of smart cities. To alleviate these concerns, we provide a framework for automatic, reliable, and uninterrupted cloud availability of services for the network data security of intelligent connected devices. This framework enables IDS to defend against security threats and to provide services that meet the users' Quality of Service (QoS) expectations. This study's intrusion detection solution for cloud network data from smart cities employed Spark and Waikato Environment for Knowledge Analysis (WEKA). WEKA and Spark are linked and made scalable and distributed. The Hadoop Distributed File System (HDFS) storage advantages are combined with WEKA's Knowledge flow for processing cloud network data for smart cities. Utilizing HDFS components, WEKA's machine learning algorithms receive cloud network data from smart cities. This research utilizes the wrapper-based Feature Selection (FS) approach for IDS, employing both the Pigeon Inspired Optimizer (PIO) and the Particle Swarm Optimization (PSO). For classifying the cloud network traffic of smart cities, the tree-based Stacking Ensemble Method (SEM) of J48, Random Forest (RF), and eXtreme Gradient Boosting (XGBoost) are applied. Performance evaluations of our system were conducted using the UNSW-NB15 and NSL-KDD datasets. Our technique is superior to previous works in

terms of sensitivity, specificity, precision, false positive rate (FPR), accuracy, F1 Score, and Matthews correlation coefficient (MCC).

Keywords: smart cities; cloud security; spark; pigeon-inspired optimizer; PSO; IDS

1. Introduction

The objective of smart cities is to efficiently and effectively manage factors such as increasing urbanization, power usage, preservation of natural resources, and the well-being of the civilian economy. A population is capable of utilizing and embracing modern Information and Communication Technologies (ICT) [1]. In the notion of smart cities, ICT plays a crucial part in policy creation, decision, implementation, and ultimate productive services [1]. According to the United Nations Population Fund, around 3.3 billion people—or 54% of the world's population—lived in urban areas in 2014; this figure is expected to rise to 5 billion (i.e., 66%) by 2030 [2]. If urbanization continues at this rate, it will have a severe impact on city management, security, and the environment. In order to effectively manage data analysis, data communications, and the successful execution of complicated strategies to maintain the smooth and secure functioning of a smart city, the efficient use of ICTs is very important [3−5].

Research has mostly concentrated on investigating potential applications and their effects on smart citizens and smart cities [6,7]. Before the recent, unexpected and widespread distributed denial of services (DDoS) attacks and ransomware threats [8,9], security and privacy in smart city systems were not considered to be critical factors. For instance, in the case of smart vehicles, a Jeep Cherokee was hacked on a highway, which prompted Chrysler to issue a recall for 1.4 million vehicles. Examples of such assaults have been mentioned in [10,11]. Complex cyberattack vectors that might affect cloud services including infrastructure, applications, and platforms continue to be a severe danger. Cyberattacks such as DDoS attacks, ransomware, and botnet attacks are frequent attempts to access cloud services and interfere with their processing resources [12].

These modifications have sparked a new wave of research into cybersecurity and data privacy in cloud computing, the Internet of Things (IoT), and intelligent city communities. Businesses have started marketing safe smart city goods [6]. The IoT can gain from improved efficiency, performance, and payload in cloud infrastructure. The development of an industrial electronic business also benefitted from Cloud Computing. Hence, IoT and cloud are extremely connected to upcoming internet technologies that are compatible with IoT systems [13]. The most accurate and effective course of action might be difficult to determine in the midst of large and complicated volumes of data.

To make the best judgment possible, modern techniques such as Artificial Intelligence (AI) and Machine Learning (ML) may be used to analyze large amounts of data [14,15]. As seen in Figure 1, the concepts of "smart cities", "big data", "data security", and the usage of AI and ML in many contexts are still in the early phases of development and will likely bring more opportunities in the future. Each element in Figure 2, which shows the architecture of a smart city application, requests security and privacy assurance procedures to address consumers' increased awareness of smart city cybersecurity [16].

The IoT is a new communication paradigm that has emerged as a result of the tremendous increase in connectedness between people, devices, and services during the past ten years. In the

upcoming generation of sustainable smart cities, this paradigm is anticipated to play a large role on the internet- and service-centric computing inside the networks of the current generation (4G/5G) and the future generation (6G and beyond) [17]. Recently, wireless technologies have emerged as a significant enabler, linking people to physical items through phones, tablets, and personal computer interfaces, which have contributed to this astonishing expansion of linked things [18]. We expect wireless transmissions to represent 2/3 of all internet traffic until 2020, with cellular/Wi-Fi connections contributing 66% of all internet protocol (IP) data [19]. As the Cloud network for data exchange grows, there is a significant risk of misusing cloud services when focusing on wireless edge devices, where sensitive and frequently semi-critical data can be fraudulently acquired, as depicted in Figure 3. By exploiting the weaknesses of wireless networks, a significant number of attackers or offenders attempted to either steal the personal information of target users or seek to obtain unauthorized access to the target's resources or applications [17].

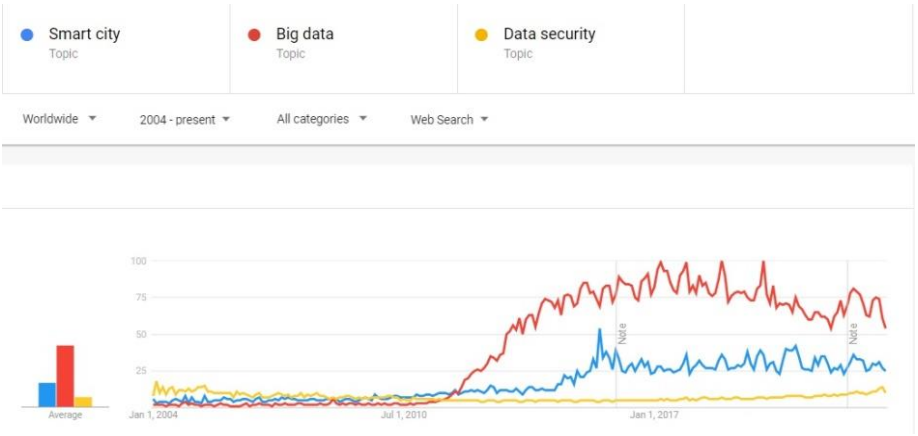


Figure 1. The popularity of big data, data security, and smart cities (Google Trends).



Figure 2. A high-level illustration of the architecture of smart city applications [16].

A cybersecurity system often consists of both networks and computer security technologies. To intercept cyberattacks, several elements (such as firewalls) and cryptographic techniques are installed, and an IDS is employed to stop external intrusions, respectively [20]. Additionally, IDS is used to define, assess, and identify unauthorized system actions, such as unauthorized access, modifications, and damage [21,22].

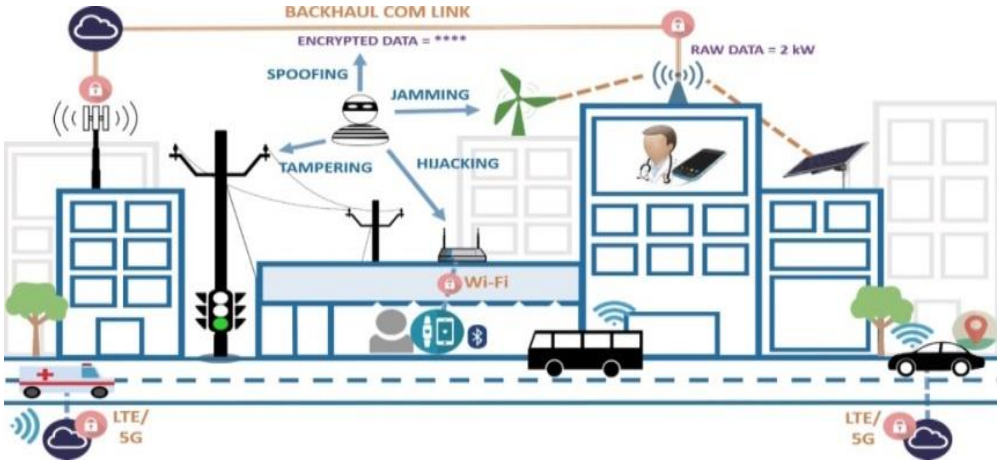


Figure 3. The complexity, scale, and extent of smart services provide more opportunities for the adversary [6].

In order to preserve the cloud network of smart cities, IDS is a security protection technique that is used to find suspicious activity in the system and quickly intercept the attacking source [23]. Depending on the types of cyber data that are accessible, IDS can be separated into host- and network-based detection. Host-based detection refers to monitoring internal resources such as logs, disc resources, and file systems on electronic devices like smartphones and laptops [21]. Antiviruses are a prime example of host-based detection. Network-based detection occurs by examining the network traffic between electronic devices and the internet [24]. In this study, we concentrate on network-based IDS for tracking fraudulent activity in the cloud network of connected device-based smart cities [24]. An effective network-based IDS should be able to identify a variety of intrusions on a cloud network of smart cities, including injection, flooding, and impersonation attacks that can originate from both internal and external attackers [25].

A smart city should include linked sensors, actuators, and relays that are safe, secure, and dependable for gathering, processing, and transmitting data in order to guarantee reliable and effective digital services; additionally, it is necessary to address the cybersecurity challenges posed by the interconnection of various devices [26]. The majority of the data is produced by IoT devices that are cloud-based and play a key part in many smart city applications [27]. IDS is frequently implemented using a centralized design, in which a single central unit is entirely in charge of evaluating all network traffic data and detecting if attacks have occurred [28]. The dependence on a single processing unit makes this strategy insecure due to a single point of failure [17].

The choice of features is crucial for creating ML models and is one of the essential steps in creating efficient IDS. The process of choosing the most key features that go into creating a strong model is known as Feature Selection (FS) [29]. FS can be performed either manually or with the aid of several methods and algorithms. Eliminating features that raise false alerts and decrease system accuracy is a crucial step in developing strong IDS [30]. As mentioned by [31], significant features convey crucial information that considerably aids in the classification process. The fact that IDS's FS decreases storage requirements, lowers processing costs, and improves test data comprehension is an important consideration. Since FS is a machine-learning topic, several different techniques are used to accomplish it. It is noted that several methods, such as the employment of intelligence patterns, swarm intelligence, Artificial Neural Networks (ANNs), deterministic algorithms, and fuzzy and rough sets, can be used to

determine features [32]. Due to their high degree of accuracy, metaheuristic algorithms are frequently utilized for FS in IDSs [33]. In this area, swarm intelligence is a key method employed in the construction and classification of metaheuristic algorithms. The AI known as "swarm intelligence" is modeled after the collective behavior of insects and swarms. It is employed to resolve challenging issues. For this research, PIO [34] and PSO [35] are two methods utilized in swarm intelligence.

A dataset that keeps expanding to the point that it is challenging to handle using traditional database concepts and technologies is referred to as big data. The data do not suit the structures of traditional database systems, are too big, move too quickly, or a combination of all three. To be beneficial, there is a need to select an alternative method of processing this data [36]. Big data is extremely challenging to store and process [37]. Hadoop is mostly used to process huge data. Hadoop uses the MapReduce framework to process the data and Hadoop Distributed File System (HDFS) to store it efficiently [38]. Spark is a framework distinguished for its responsiveness. It tries to speed up batch workloads by performing the entire calculation in memory and processing optimization [39]. Additionally, spark is effective at performing iterative calculations, making it a good choice for the creation of large-scale machine-learning systems [40]. Nowadays, organizations are placing more emphasis on big data systems to manage such data and to use them for commercial expansion. The top seven data drivers include instances of big data from the financial, Internet, mobile, Smart cities-based data, science, sensor, and streaming industries.

The use of ML approaches to enhance IDS performance has been the subject of many studies [41,42]. However, these systems are limited because they use only one classifier, which prevents them from identifying and thwarting serious threats. In order to address this problem, Hansen and Salamon [43] have created several ensemble algorithms that outperform single classifiers. In this context, Pham et al. [44] and Rashid et al. [45] highlight numerous factors that might affect IDS performance, including feature selection, base classifiers, and ensemble techniques. This suggests the value of investigating ensemble-based IDS from diverse angles. Tree-based ML techniques have demonstrated promising results in predictive analytics [46]. Additionally, compared to many other kinds of ML models, tree-based classifiers may be trained significantly more quickly [47]. In this paper, we provide a tree-based SEM for enhancing IDS performance after taking their applicability into account.

1.1. Motivation

To provide connected solutions for the general public, smart cities combine the IoT with a range of software, user interfaces, and communication networks. The most likely security threats to the networks of smart cities include various malicious attacks like DDoS, ransomware, remote recording, routing attacks, and data leakage attacks [6]. A DDoS attack is when many compromised machines launch a denial of service (DoS) attack on the smart cities network. With this attack, the attacker floods the server with erroneous request packets to block it, thereby depriving users of the resources and lengthening response times. Hence, it is important to make sure that smart cities are protected from cyberattacks, hacking, and data theft. The network of smart cities is protected by several security measures, including access control algorithms, public key-based security methods, and IDSs. However, there are no effective security options. An efficient IDS is necessary to provide integrated solutions that meet the fundamental security goals of availability, integrity, confidentiality, and accountability of the smart city network.

Currently, intrusion detection in smart cities has drawn increased interest in the cybersecurity field. Numerous FS and intrusion detection algorithms have been studied during the last few decades. Due to limitations, in many cases, these algorithms could no longer address the difficulties in real-time and distributed a nature of applications. The application of population-based metaheuristic algorithms in optimization, including feature selection, has recently increased. In order to handle the cloud network data from smart cities, the IDS must be scalable and dispersed. The use of only one classifier makes it impossible to detect and neutralize severe threats. Tree-based ML techniques have demonstrated promising outcomes in predictive analytics. In order to reduce potential future security attacks against the cloud network data of smart cities, the motivation is to propose an IDS using tree-based SEM ML classifiers. Additionally, we adopt the HDFS and Spark framework in our proposed work because the data generated from the cloud network of smart cities is enormous and can only be stored and processed in a cloud environment.

1.2. Contribution

IDSs are crucial in preventing intrusions into the cloud network of smart cities, which contains private data and information. This research lists the most recent developments in FS and classification techniques for IDSs using cloud network data from smart cities. The following tasks are included: using a wrapper-based FS approach that uses PIO and PSO, storing the data in the HDFS, processing cloud network data for smart cities using WEKA's Knowledge flow [48], and classifying the cloud network traffic of smart cities using a tree-based SEM approach that uses J48 [49], RF [50], and XGBoost [51]. The UNSW-NB15 [52] and NSL-KDD [53] datasets were used to assess the system's performance. HDFS is used for storing and processing massive datasets. WEKA is used to train ML for massive data mining and analysis. The "DistributedWekaSpark" package allows us to configure Spark and WEKA together.

This effort's core contribution consists of examining the breadth and scope of the effects of cyberattacks, and the results are frequently utilized as the foundation for new legislation and regulations. Computer scientists and researchers make up the second branch. They examine the technological tools to ensure they adhere to the security and privacy standards set out by smart city rules. The research community has recognized that the technological implementation of smart city security is also a complicated issue, with the system's total security being determined by the most delicate link in the network. This insight is the source of vulnerabilities in existing smart city systems since developers erroneously think they can boost their products' security by protecting one aspect of the system while ignoring others.

The main contributions are pointed out as follows:

(1) The suggested IDS has been modified to ensure that secure communication with service providers only happens when it is performed between people of smart cities, trustworthy third-party organizations, and service providers.
(2) This research effort proposes a unique approach for automatically identifying intruder attacks by classifying both legal and malicious data.
(3) The hybrid monitoring system that makes up the intrusion detection mechanism combines data preprocessing methods, PIO and PSO for data dimensionality reduction, HDFS for data storage, and a tree-based SEM technique that employs J48, RF, and XGBoost for attacks classification using WEKA's Knowledge flow and Spark framework.

(4) A model built on the Spark platform to divide the processing of cloud network data for smart cities in order to save computation time.
(5) With reference to benchmark datasets, the suggested model is assessed to demonstrate how it outperforms standard baseline methods.

1.3. Paper structure

The structure of the article is as follows. The related work is discussed in Section 2. Prerequisites for the proposed methodology, which include dataset enumerations, FS tools (PSO and PIO), classification approaches (J48, RF and XGBoost), and WEKA and Spark frameworks are discussed in Section 3. In Section 4, the suggested tree-based SEM ML algorithms implementation on Spark and WEKA are used to evaluate cloud network data from smart cities. In Section 4, a performance assessment, model performance, implementation, and comparison of the proposed model are included. Finally, Section 5 concludes the paper and lays out the research agenda for the future.

2. Related work

In this section, we review related research and present the most pertinent related work to address barriers that need to be considered in this study. Without question, the emergence of smart cities will improve many aspects of urban life, while significantly increasing vulnerability. We research these concerns on a variety of levels. The associated works are divided into many categories, including secure Cloud/Edge/Fog/IoT ecosystems for smart cities, IDSs for smart cities, and FS techniques. The following subsections explain these categories in more depth.

2.1. Secure Cloud/Edge/Fog/IoT ecosystem for smart cities

The architecture called Hybrid IoT was proposed by Qian et al. [54] to enable the well-organized transmission, caching, and computation of massive data produced by widely scattered and substantial IoT devices that are installed in smart cities. A paradigm for monitoring in the transportation sector was proposed by Garg et al. [55]. They used real-time analytics and apps at several levels to solve the issue of security risks in aerial vehicles. To discover cyber-threats in smart cars, they applied the probabilistic data structure method. By gathering data from moving objects and passing the load to edge devices through aggregators, extreme aerial vehicles serve as data providers. The real-time security of the odd vehicle movements was assured by the creators. In order to describe the function of cloud computing in providing storage, computation, databases, and a variety of application services for access through the internet, Dener [56] reviewed many research publications. These cloud-based services facilitate information sharing and integration amongst various smart city systems.

Applications based on reinforcement learning have produced successful outcomes in mobile edge computing (MEC), vehicular edge computing (VEC), and other areas. Reinforcement learning (RL) has emerged as an important class of algorithms in AI. The AI discipline of RL contributes to the further optimization of energy usage and the performance of MEC [57]. In a network, MEC provides execution resources, such as storage and calculations, which are close to the users and can be used to process and store content, as well as deliver services. Two sub-problems are created from

the energy consumption optimization problem: computation optimization (data segmentation) and transmission optimization (time division). The authors [57] have suggested DDQNL-IST, an intelligent game method that combines distributed LSTM (Long Short-Term Memory) and DDQN (Double Deep Q-Network) with an intermediate state transition (IST). The results of the experiments demonstrate that the suggested DDQNL-IST can perform better in terms of average latency and energy cost. Due to the restricted processing resources available to the VEC servers, the authors of the research [58] suggest a resource management strategy based on Deep-RL (DRL) to motivate the VEC servers.

An innovative framework known as VCoT, presented by Khattak et al. [59], merges IoT and vehicular-networking clouds. The purpose of IoT-VC (VCoT) for numerous real-world applications, including smart traffic signals, home automation, and smart cities, is thoroughly explained in the article, along with the challenges that must be overcome. For the progress of smart cities, Kaur et al. [60] suggested an architecture that relies on cloud computing and IoT. The author concentrated on the various cloud characteristics and utilized IoT to deploy them to improve smart cities. The author used the smart city of Dubai as a case study and suggested a scenario-based design for healthcare in a smart city. A big-data analysis based on a smart city employing cloud computing infrastructure was described by Massobrio et al. [61]. The Hadoop framework was used to implement the map reduced parallel model. In essence, they concentrated on two cases: the estimate of the origin-to-destination matrix and public transportation services. The first case uses information about prior sites, whereas the latter case uses information about ticket sales. The actual outcome demonstrates how well the model supports a large volume of data.

PROTeCt (Privacy aRchitecture for IntegratiOn of Internet of Things and Cloud computing), is a device-based security system that enhances user privacy through a cryptography-based system in which only interested users may access their data, which is stored on a cloud in encrypted form to guard against unauthorized access [62]. The gateway was authenticated in this operation. Users must repeatedly register and accept invitations to join the network, which raises the cost of communication. A framework for effective IDS that is more suitable for IoT-based applications was provided by the authors in [63]. To deal with security anomalies for IoT networks, the suggested framework uses machine learning-based methodologies. The link measurement required to choose the best routing option is missing from the research. These limitations increase the percentage of route breakages and the rate of packet loss.

The authors in [6] investigated the deployments of cyber-physical systems in smart cities. They focused on the security issues associated with smart infrastructure and the impact that ransomware causes on governmental organizations, the healthcare system, and transportation. Moreover, the solutions include game theory, IDS, and cryptography. It is important to note that the writers also focused on human error. The authors in [17] introduced an IDS based on ML in a semi-distributed and distributed mode for the resource-constrained IoT. This IDS is based on different FS strategies, and each technique is investigated independently.

The authors of [64] focused on the security of IoT devices in the smart city and presented an architecture dubbed the Anomaly Detection-IoT (AD-IoT) system that was built on RF ML. The authors of [65] presented a novel DRL-based architecture to protect a smart city's digital infrastructure from any kind of cyber incursion and for early detection of intrusions based on data behavior. To reduce latency and energy use, the authors of [66] introduced a neuro-fuzzy-based secure PSO computational offloading system for the Fog-Cloud-IoT context. The authors of [67]

envisioned an ML-based secure cloud service for connected automobiles that would identify cyberattacks and satisfy user QoS and Quality of Experience (QoE) requirements.

Table 1. Brief literature Survey for implementing secure Cloud/Edge/Fog/IoT ecosystem for smart cities.

Reference	Year	Method Employed	Summary
[60]	2016	Cloud and IoT	The integration of any smart city application requires cloud computing and IoT.
[65]	2017	Restricted Boltzmann Machine	Generalized method to spot and stop DoS in smart cities.
[63]	2017	ML based IDS	A cutting-edge intrusion detection technology that recognizes security irregularities in IoT networks using ML algorithms
[61]	2018	Hadoop	Smart city Big Data analysis paradigm utilizing cloud computing infrastructures.
[62]	2018	PROTeCt	User privacy is increased through the privacy architecture for the IoT and CC integration.
[55]	2018	Triple Bloom filter PDS	A data-driven transportation optimization model that uses a PDS-based method to identify cyber threats in smart cars.
[54]	2019	UDN and MEC	Lower the energy use and end-to-end latency of computing data from large IoT devices installed in a smart city.
[56]	2019	Cloud Computing	Cloud-based services facilitate information sharing and integration amongst various smart city systems.
[59]	2019	VCoT	A new framework for architectural and communication design that successfully integrates IoT and cloud-based vehicular networking.
[6]	2019	Reviews for Security for Smart cities	A multi-level structure, including a “security level”, can be used to conceptualize smart cities. All other levels’ weaknesses are caused by the security level.
[64]	2019	RF-based binary classification	AD-IoT system is suggested as a solution to deal with the cybersecurity concerns posed by IoT in smart cities.
[66]	2019	Neuro-Fuzzy Model and PSO	SecOFF-FCIoT
[67]	2019	Deep belief and DT	IDS offers services that satisfy users’ QoS and QoE needs.
[17]	2020	SAE and MLP	IDS is based on semi-distributed and fully-distributed approaches.
[57]	2021	Energy consumption optimization in MEC.	RL-based performance enhancement of MEC by energy consumption optimization.
[58]	2022	DRL and Stackelberg game-based resource management for VEC.	A DRL-based resource management plan is suggested to boost vehicle and VEC server revenues.

UDN- Ultra-Dense Networking, PDS- probabilistic data structure, VCoT- vehicular networking clouds with IoT, SAE-Stacked Auto-Encoder, MLP- Multi-layer perceptron, SecOFF-FCIoT- Secure offloading in a Fog-Cloud-IoT.

The aforementioned papers offer the first step towards the ecosystem of smart cities. As the number of connected devices grows due to technological innovation, it becomes impractical to individually connect many low-end 'things' to management or analytics systems. In general, some end-system devices produce so much data that the core communication channel may become saturated. An array of crucial tasks, including data filtering and processing, and security are carried out by the Cloud/Edge/Fog/IoT ecosystem for smart cities. Security concerns in smart cities are application-based. For instance, the security flaw in smart meters might result in energy disruptions and ineffective Smart Grids. Therefore, more sophisticated and cutting-edge techniques based on big data analytics are required to guarantee the cyber safety and security of smart city applications. However, the aforementioned cryptographic and technical approach is viewed as insufficient due to the increasingly sophisticated and varied cyber-attacks. This drives the creation of an IDS with built-in intelligence to move away from "one-shot" security protection and to include a sophisticated method of continuous learning from changing network data. The discussion and summary of existing works are included in Table 1.

2.2. IDS for smart cities

IDSs have long been a topic of study in the field of communication networks; however, the practical need for such networks has only recently led to a shift in emphasis to IDS for smart city-based networks. IDSs are divided into host-based [68,69] and network-based [70,71] subcategories. Due to several factors, host-based IDSs are considerably less effective in identifying corrupted IoT devices. First, the development of detection algorithms uses the energy and computing of smart IoT devices. Second, certain IoT devices have restrictions on what software may be installed on them. Third, deploying detection techniques on many diverse IoT devices linked to smart networks in smart cities is extremely difficult. Lastly, IoT device makers rarely incorporate detection techniques into their designs [72].

DeepCoin is an innovative Deep Learning (DL) and blockchain-based energy framework for smart grids [73]. The blockchain-based system is divided into five stages: setup, agreement, block creation and consensus, view modification, and conclusion. It provides a high throughput methodology and contains a revolutionary, trustworthy peer-to-peer energy system based on the useful Byzantine fault tolerance algorithm [73]. The suggested system generates blocks using hash functions and short signatures to guard against smart grid attacks. It is suggested to use the statistical correlation between measurements for unsupervised anomaly identification with the objective to create a scalable anomaly detection engine that is appropriate for large-scale smart grids that can distinguish between a genuine malfunction and either a disturbance or sophisticated cyber-attack [74] The suggested approach uses feature extraction while learning about the causal relationships between the subsystems using symbolic dynamic filtering (SDF), which also helps to lighten the computing load.

The authors in [75] presented an intrusion detection/prevention system (IDPS) with fog-assisted software-defined networking (SDN) using an enhanced decentralized computing structure known as fog-computing as an IoT framework. To address an IoT scalability issue, they suggested a useful technique for allocating fog resources. Additionally, they examined four classifiers to identify intrusions and provided design recommendations to control cybersecurity threats at the edge of the IoT network and to spot anomalies. An online Sequential Extreme Learning Machine (OS-ELM) was used by the authors of [76] to construct a fog-oriented IDS, which summarizes the identified intrusion. First, fog-nodes will identify the malicious traffic from the IoT environment, and the information of the

identified intrusion is further transferred to the cloud server. Since the suggested approach is a distributed IDS, it can offer scalability, interoperability, and flexibility. However, it is unable to summarize the data without a cloud server and will become problematic once the cloud server goes down.

A supervised IDS was suggested by the authors of [77] for an IoT network in a smart house. The three-layered model of the proposed IDS architecture worked to identify malicious packets. Three experiments were tested over the layers using nine classifiers. Consequently, the J48 classifier produced F-measure values of 96.2%, 90%, and 98% for each of the three experiments, respectively, to obtain the best performance. The system's drawback is that it must integrate the three levels in order to detect malicious communications. The entire system will have problems if one of the layers fails. To assist managers of smart cities in defining the most sensitive threats, researchers in [78] suggested an intrusion detection framework and an attack categorization scheme. Additionally, this paper demonstrates how a One-Class Support Vector machine (OC-SVM) and rule-based detection may be used together to dramatically enhance detection results.

In [79], the authors investigated the viability of using single model classifiers in an ensemble learning setting to identify cyberattacks in IoT-based smart city applications. The tests using the most recent IoT attack the database, stacking a component of the ensemble technique outperforms single models in distinguishing attacks from benign samples. In terms of various performance emetrics, information gain (IG) is employed for FS and classification outcomes outperform either single or other ensemble models. Diro and Chilamkurti suggested a DL model to detect distributed intrusions in a social IoT network using the NSL-KDD open-source dataset, which logs attack data in both distributed and centralized systems; their model achieved 99.2% and 98.27% accuracy for binary-class and multi-class identification, respectively [78].

Due to the negative effects of low-frequency threats, such as user to-root (U2R) and remote-to-local (R2L) attacks, Pajouh et al. introduced a two-stage dimension reduction and classification approach to identify anomalies in IoT backbone networks [18]. After reducing the dataset's features using principal component analysis (PCA) and linear discriminate analysis (LDA), they employed naive bayes and K-Nearest Neighbor (KNN) to find anomalies. This method resulted in an identification rate of 84.82% [18]. However, this method is centralized and was only tested for DoS, remote-to-local, user-to-root, and Probe attacks. In [80], Kozik et al. presented an attack detection system that used the Apache Spark cloud architecture and the ELM method. The accuracy levels of this investigation, which focused on the three key IoT system, used cases of scanning, command and control, and infected host were 99%, 76%, and 95%, respectively.

The Gain Ratio (GR) FS approach, based on ANN and Bayesian networks, was suggested [81] and the performance was assessed on the KDD'99 and NSL-KDD datasets, with ensemble techniques achieving 99.42 and 98.07% accuracy, respectively. An ensemble technique that incorporates Naive Bayes, Bayesian Net, and DT classifier was put out by Haq et al. [82], while using FS methods such as Best First Search, Genetic, and Rank Search, where they were able to extract the common features. The ensemble methodology generated a 98% true positive rate when examined using the 10-fold cross-validation approach. A reduced error pruning tree (REPTree) was utilized as the basis classifier in the bagging ensemble approach that Gaikwad et al. [83] developed on the NSL-KDD dataset; their model had an accuracy rate of 81.29%. Jabbar et al. [84] suggested an ensemble approach that combined an Alternating DT (ADTree) with KNN, and the performance evaluation showed that the proposed ensemble outperformed the current strategies in terms of the Detection Rate (DR) (99.8%).

Zhou et al. proposed a FS and ensemble method-based IDS model in [47], where an optimal FS

was achieved using a combination of Correlation-based FS (CFS) and Bat algorithm, followed by an ensemble method made up of the Decision Tree (DT), RF, and Forest by Penalizing Attributes (Forest PA) algorithms. A hybrid IDS that combined the C5 classifier and OC-SVM was introduced in [85]. Using DT and RF trees as the basic classifiers, the authors of [44] developed bagging and boosting ensemble approaches. Experiments were conducted on the NSL-KDD dataset, where it was discovered that bagging with DT produces superior outcomes.

Table 2. Brief of literature Survey for implementing IDS which are suitable smart cities.

Reference	Year	Method Employed	Summary
[81]	2014	Ensemble-based Multi classification	GR FS approach combined with ANN and Bayesian Net classifiers for an IDS.
[82]	2015	Ensemble-based Multi classification	J48, Bayesian Network, and NB classification model ensemble utilizing a hybrid FS technique.
[78]	2017	OC-SVM	Rule-based detection using OC-SVM is employed to increase performance.
[84]	2017	Ensemble-based Multi classification	The ADTree and KNN oriented cluster- based ensemble classifier is constructed.
[75]	2018	RNN, MLP, and ADT	IoT Network Anomaly Detection Using a Fog-Assisted SDN
[76]	2018	OS-ELM	The testing findings indicate that the fog nodes identify attacks 25% more quickly than cloud-based implementations while maintaining a low false alarm rate.
[86]	2018	Neural Network based Multi classification	A novel method of cybersecurity called DL makes it possible to identify threats in the social IoT.
[80]	2018	ELM based Multi classification	Sharing the traffic load between edge and cloud for effective traffic classification using ELM.
[44]	2018	Ensemble-based Binary classification	Bagging ensemble model with J48 as the basic classifier.
[85]	2019	Multi-classification by C5 classifier and One class-SVM	Ensemble of C5 classifier with the OC-SVM classifier, to identify both known intrusions and zero-day attacks with high detection accuracy and low false-alarm rates.
[18]	2019	NB and kNN	Two-tier classification module to spot unusual R2L and U2R attack behaviors.
[77]	2019	J48	A three-layered IDS to identify several common network-based cyberattacks on IoT networks.
[74]	2019	Symbolic Dynamic Filtering (SDF) and Boltzmann Machine	This technique identifies unobservable intrusions in smart grids.
[47]	2020	Ensemble-based Multi classification	Combination of the C4.5, RF, and Forest by Penalizing Attributes (Forest PA) algorithms for attack identification through the voting process.
[79]	2020	ANN, SVM LR, DT, RF, and KNN	Examine a machine learning-based attack and anomaly detection strategy to counter and reduce IoT cybersecurity vulnerabilities in a smart city.
[73]	2020	RNN, Blockchain, and Byzantine fault tolerance algorithm.	IDS to identify network attacks and fraudulent energy network transactions.
[83]	2021	Ensemble-based Binary classification	Utilizing REPTree as the foundation class, the ensemble's bagging method is utilized to construct an IDS.

The aforementioned models were created for IoT networks; however, because they primarily focused on the network structure, they did not take the resource limits and limitations that exist within IoT networks into account. The models' findings indicated that impersonation attacks appear to have a less favorable outcome or detection rate. However, these security measures come at a high performance cost and are inappropriate for the accepted smart city context. The discussion and summary of existing works are comprised in Table 2.

2.3. Feature selection procedures

The important step in intrusion detection in cloud Internet traffic data of smart cities is feature selection. It is difficult to categorize and detect anomalous and unknown classes without feature extraction and selection. Thus, for the field of network traffic identification, a comprehension of FS and extraction is very essential. There are many intrusion detection techniques [87−94] employed to provide security for communication networks while employing the NSL-KDD and UNSW-NB15 datasets.

In [95], Tama et al. introduced an ensemble-based IDS, which combined a two-stage classifier with a hybrid FS technique including PSO, ant colony algorithm (ACO), and genetic algorithm (GA), applied on UNSW-NB15 and NSL-KDD datasets. By altering the value of the parameter n, which stands for the number of particles, population size, and ants in PSO, GA, and ACO, respectively, experiments were conducted to ascertain the ideal configuration for feature selection. With an accuracy of 99.5570 ± 0.134%, PSO with $n = 2$ clearly revealed the best classification result. This case produced a collection of 37 features. The feature set of 19 was produced using PSO with $n = 5$, which yield the highest classification accuracy of 97.0550 ± 0.125% for UNSW-NB15.

An IDS that is based on stacking ensembles was proposed by Smitha et al. [96]. Experiments were conducted on the heterogeneous datasets UNSW-NB15 and UGR'16. Only the best features were retrieved after the most important traits were provided weights in order to prioritize them. The IG-based hashing approach was used to minimize the dimension of the features; only 11 characteristics from the UNSW NB-15 dataset were chosen [96]. Alternatively, five characteristics of the UGR'16 dataset's were taken into consideration. In order to create a hybrid IDS, Salo et al. [97] combined an ensemble approach with two FS techniques IG-PCA. The IG-PCA ensemble approach selected seven features out of 20 features of the ISC2012 dataset and achieved the highest accuracy of 99.011. For the NSL-KDD dataset, 12 features out of 41 features were selected by the IG-PCA ensemble approach to gain the highest accuracy (98.24%). Twelve out of 24 features were selected by the IG-PCA ensemble approach using the Kyoto 2006+ dataset to obtain the highest accuracy of 98.95.

Zhou et al. introduced an ensemble-based, feature-selected IDS in [47]. They coupled the Bat algorithm with CFS (CFS-BA) to choose features. After choosing the features, they conducted the experiment using the NSL-KDD, AWID, and CICIDS2017 datasets. The CFS-BA approach selected 10 out of 41 features of the NSL-KDD dataset. For the AWID dataset, eight out of 155 features were selected. Alternatively, 13 out of 80 features were selected by the CFA-BA approach using the CICIDS2017 dataset. In [79], Rashid et al. examined several ML techniques to identify cyber-attacks from IoT-based smart city applications using an FS approach. For UNSW-NB15 and CICIDS2017, the authors employed the information-gain model and chose 25 features. To improve the suggested ensemble approach, they added FS algorithms to select the most pertinent features. The top 20 out of 41 features from the NSL-KDD dataset were chosen using the proposed model and the SelectKbest FS algorithm [45]. The results of FS utilizing CFS as the

feature evaluator and PSO as the search strategy (CFS+PSO) were presented in this work [98]. For PSO, we assume that there are 50 particles, that the inertia weight constant is 0.33, and that the values of c_1 and c_2 are equal at 0.34. After utilizing CFS+PSO to execute FS on the NSL-KDD dataset, 11 important features are effectively acquired.

To select the best subset of attributes for IDS, a unique combination technique based on the Iterative Dichotomiser 3 (ID3) algorithm and the bee algorithm (BA) was proposed [97]. The BA was used to offer a subset of features, while the ID3 approach acted as a classifier. When applied to the KDD Cup 99 dataset, the results showed that the proposed model performed better for DR and Arrival Rate (AR) when the number of features was less than 30. An IDS based on FS and a clustering utilizing filter and wrapper approaches were proposed in study [31]; this featured grouping based on a linear correlation coefficient (FGLCC) algorithm and a cuttlefish algorithm (CFA), which are the names of the filter and wrapper approaches, respectively. The suggested technique used a DT as the classifier. Using the KDD Cup 99 dataset, FGLCC and FGLCC-CFA FS techniques choose 15 and 10 features, respectively [31].

Table 3. Brief of literature survey for implementing feature selection.

Reference	Year	Method		Dataset	Total Features	Feature Selected
[102]	2015	ID3-BA		KDD CUP 99	41	<30
[98]	2017	CFS-PSO		NSL-KDD	41	11
[99]	2018	IWD		KDD CUP 99	41	9
[100]	2019	Firefly		KDD CUP 99	41	10
[31]	2019	FGLCC		KDD CUP 99	41	15
		FGLCC-CFA		KDD CUP 99	41	10
[97]	2019	IG-PCA		NSL-KDD	41	12
				Kyoto 2006+	24	12
[95]	2019	PSO		NSL-KDD	41	37
				UNSW-NB15	49	19
[96]	2020	IG-Hashing		UNSW-NB15	49	11
				UGR'16	12	5
[47]	2020	CFS-BA		NSL-KDD	41	10
				AWID	155	8
				CIC-IDS-2017	80	13
[79]	2020	IG		UNSW-NB15	49	25
				CIC-IDS-2017	80	25
[101]	2020	PIO	SPIO	NSL-KDD	41	18
				UNSW-NB15	49	14
			CPIO	NSL-KDD	41	5
				UNSW-NB15	49	5
[45]	2022	SelectKbest		NSL-KDD	41	20

Acharya and Singh [99] developed a unique method for choosing IDS features that used the Intelligence Water Drops (IWD) algorithm. IWD is also recognized as a metaheuristic-based swarm intelligence optimization technique. With the help of the KDD CUP99 dataset, this strategy was assessed. The suggested wrapper model obtained a minimum of nine features out of a total of 41. The suggested study used a filter and wrapper-based technique using the firefly algorithm in the wrapper to choose the

features from the KDD-CUP99 dataset since FS affects how quickly the analysis is completed [100]. Ten features are considered. This method employs the selection process by means of an optimizer that was inspired by pigeons. The conventional approach for binarizing continuous swarm intelligence algorithms is contrasted with a novel approach for a continuous PIO [101]. The UNSW-NB15 dataset has 49 features, whereas the KDDCUP 99 and NSL-KDD datasets each have 41. However, not all of these features are crucial for creating IDS. Both the Sigmoid PIO (SPIO) and the Cosine PIO (CPIO) binarized versions of PIO for FS chose 10 and 7 features from the KDD-Cup99 dataset, respectively. From the NSL-KDD dataset, the PIO for FS (SPIO and CPIO) chose 18 and 5 features, respectively. From the UNSW-NB15 dataset, the PIO for FS (SPIO and CPIO) chose 14 and 5 features, respectively.

It is crucial to note that while most of the described methods have been evaluated using the most recent KDD dataset (i.e., NSL- KDD and UNSW-NB15), a small number have also been tested using the extremely old KDD'99 dataset (1999). In this research, we used both datasets to evaluate the proposed framework. In conclusion, it should be noted that the research does not agree on a specific number of features or subset of features, and that the proposed FS algorithms deal with a tradeoff between the accuracy and FPR. The brief of existing works is comprised in Table 3.

3. Prerequisites

3.1. Dataset enumeration

3.1.1. NSL-KDD

The KDD Cup 99 dataset has been revised to become NSL-KDD. The dataset is comprised of 41 attributes that are divided into 5 classes—4 attack groups and 1 normal class—that are explained in more detail in Table 4. The 42nd attribute, aka class attribute, gives details about these groups and has either positive or negative examples. Here, we outline the most common forms of malicious behavior, which are divided into 4 categories of attacks.

- *DoS attacks*- Attacks that restrict the services of legitimate users fall under the term DoS. A few examples include Smurf, teardrop, SYN flooding, and Neptune.
- The term "*User to Root*" (U2R) refers to situations in which an attacker takes control of local machines by abusing flaws within them. U2R is a type of exploit in which the attacker first gains access to a regular user account on the system, and then uses that account to exploit a security hole to take control of it [103]. A few examples include rootkit, espionage, buffer overflow, and SQL attacks.
- *R2L (Root to Local)*- R2L occurs when an attacker, who can send packets to a system across a network but who lacks an account on that machine, uses a vulnerability to acquire local access as a user of that machine [103]. A few examples include Warezmaster, Imap, multihope, and spy.
- A computer network is probed to learn more about it with the apparent goal of getting beyond security measures. An example of a *probe attack* is when the attacker uses a traffic analysis to learn more about the network. Examples include nmap, satan, ping-sweep, and port-scan.

The training part of the NSL-KDD dataset (KDDTrain+) is a dataset of 125,973 records, of which 67,343 records are normal and the remaining 58,630 records are anomalous. The test part of the dataset (KDDTest+) is comprised of 22,544 records. of which 9711 records are normal and the 12,833 records are anomalous.

3.1.2. *UNSW-NB15*

The dataset is critical for evaluating and measuring the performance of IDS. During the past few decades, IDS datasets have been introduced. Moustafa et al. [52] generated the UNSW-NB15 dataset lately. The UNSW-NB15 testbed is shown in Figure 4. The UNSW-NB15 dataset is a combination of a real-world network operation and a synthetically modified attack. In this study, the UNSW-NB15 dataset is used. IXIA PerfectStorm, which is an attack creation tool, was used to produce the UNSW-NB15 dataset. It includes both modified and actual attacks from nine different families. Different servers are targeted in these attacks. At the beginning of 2015, the authors acquired tcpdump traces of the network traffic for a total of 31 hours. In addition, the dataset consists of 12 methods that are utilized to provide 49 features for the class label [96]. For each network flow, these network records were utilized to create a dataset containing 49 features. Some of the features are numerical, while others are statistical. Other qualities refer to the values of time stamps.

There are 175,341 records in the training set and 82,332 records in the testing dataset, which include all types of attacks, as well as typical traffic samples. There are 45 features in the testing and training datasets. These features are listed in Table 5. The UNSW-NB15 dataset has been subjected to nine different types of attacks.

Table 4. NSL-KDD dataset with feature number, name, and type.

Feat. No.	Name	Type	Feat. No.	Name	Type
42	Class	Nominal	21	Is host login	Discrete
41	Destination host service rerror rate	Discrete	20	Number of outbound cmds	Discrete
40	Destination host rerror rate	Discrete	19	Number of access files	Discrete
39	Destination host service serror rate	Discrete	18	Number of shells	Discrete
38	Destination host serror rate	Discrete	17	Number of file creations	Discrete
37	Destination host service different host rate	Discrete	16	Number of root	Discrete
36	Destination host same source port rate	Discrete	15	Su attempted	Discrete
35	Destination host different service rate	Discrete	14	Root shell	Discrete
34	Destination host same service rate	Discrete	13	Number of compromised	Discrete
33	Destination host service count	Discrete	12	Logged in	Discrete
32	Destination host count	Discrete	11	Number of failed logins	Discrete
31	Service different host rate	Discrete	10	Hot	Discrete
30	Different service rate	Discrete	9	Urgent	Discrete
29	Same service rate	Discrete	8	Wrong fragment	Continuous
28	Service rerror rate	Discrete	7	Land	Discrete
27	Rerror rate	Discrete	6	Destination byte	Continuous
26	Service serror rate	Discrete	5	Source byte	Continuous
25	Serror rate	Discrete	4	Flag	Discrete
24	Service count	Continuous	3	Service	Discrete
23	Count	Discrete	2	Protocol type	Discrete
22	Is guest login	Discrete	1	Duration	Continuous

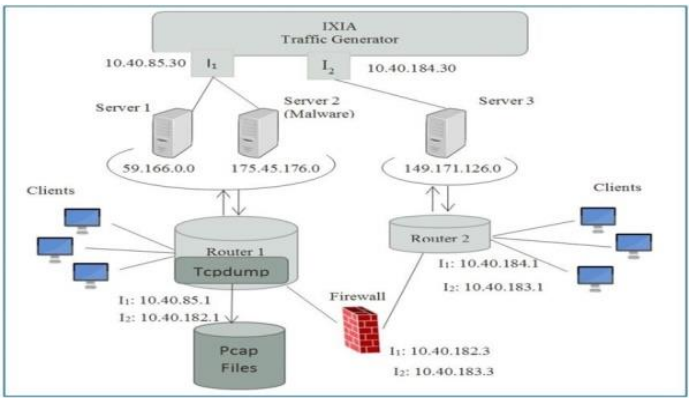


Figure 4. The Testbed Visualization for UNSW-NB15 [52].

Table 5. Enumeration of the employed dataset.

Feat. No.	Feat. Name	Feat. No.	Feat. Name	Feat. No.	Feat. Name
45	label	30	transdepth	15	sloss
44	attack_cat	29	dmean	14	dload
43	issmipsports	28	smean	13	sload
42	ctsrvdst	27	ackdat	12	dttl
41	ctsrcltm	26	synack	11	sttl
40	ctflwhttpmthd	25	tcprtt	10	rate
39	ctftpcmd	24	dwin	9	dbytes
38	isftplogin	23	dtcpb	8	sbytes
37	ctdstsrcltm	22	stcpb	7	dpkts
36	ctdstsportltm	21	swin	6	spkts
35	ctsrcdportltm	20	dit	5	state
34	ctdstltm	19	sjit	4	service
33	ctstatettl	18	dinpkt	3	proto
32	ctsrvsrc	17	sinpkt	2	dur
31	responsebodylen	16	dloss	1	id

The training section of the UNSW-NB15 dataset consists of 175,341 records, of which 56,000 are normal and 119,341 are attacks. The test part of the dataset is comprised of 82,332 records, of which 37,000 records are normal and 45,332 records are anomalous.

3.2. Feature selection tools

3.2.1. *Pigeon Inspired Optimization (PIO)*

One of the recently created bio-inspired swarm intelligence algorithms is the PIO [104]. Pigeons' homing behavior was influenced by two primary operators: landmark operators and map and compass operators. According to the research on pigeon homing abilities, the pigeon's capacity to find its way home is caused by small magnetic particles that are found in its beak (i.e., through the trigeminal nerve). These particles communicate with the brain of the species, and Pigeons can feel

the earth's magnetic field using their magneto-reception abilities. Additionally, they can utilize the sun's height as a compass to change their orientation [101]. The pigeons grow less dependent on the map and compass operator as they approach their objective [104]. Guilford and others in [105] devised the PIO algorithm, which is based on the two primary operators employed by pigeons and is meant to match their behavior.

By changing the position X_i and velocity V_i of pigeon i throughout each iteration, this operator may be mathematically stated in a more straightforward manner. Based on the value of the current iteration t in Eqs (1) and (2), the values of X_i and V_i are changed for the following $(t + 1)^{th}$ iteration, as stated in [104]:

$$V_i(t+1) = V_i(t).e^{-Rt} + random.\left(X_g - X_i(t)\right) \tag{1}$$

$$X_i(t+1) = X_i(t) + V_i(t+1), \tag{2}$$

where X_g is the global best solution, $X_i(t)$ stands for the current position of a pigeon at iteration t, and $V_i(t)$ stands for its current at iteration t. R stands for the map and compass factor. "*random*" is a uniform random number in the range [0, 1]. Equation (1) is used to calculate each pigeon's velocity in the traditional manner, and Eq (3) uses a sigmoidal function to convert the velocity into binary form:

$$S\left(V_i(t)\right) = \frac{1}{1+e^{\frac{-v_i}{2}}} \tag{3}$$

$$X(t)_{(i,p)}[i] = \begin{cases} 1, if\left(S\left(V_i(t)\right) > r\right) \\ 0, otherwise \end{cases}, \tag{4}$$

where r is an evenly distributed random number and $V_i(t)$ is the pigeon velocity in iteration t.

In this research, we apply the PIO's new binary version to an IDS FS method. The features selected by CPIO generate efficient results compared to SPIO and traditional PIO in terms of accuracy, TPR, and FPR [101]. The outcome demonstrates that the proposed CPIO, which used the cosine similarity to binarized the solution velocities rather than the sigmoid function, had a faster convergence [101]. Hence, we adopted the Cosine version of PIO for FS purposes. The cosine similarity was employed by CPIO to determine the pigeons' velocity. Initial binary values of either zero or one were chosen at random to set the value of the solution. The cosine similarity formula is used to calculate the velocity and to determine how similar the local and global pigeons, X_p and X_g, are to one another. The pigeon velocity calculation is shown in Eq (5). The position of the pigeon will be updated in accordance with Eq (6) based on the probability that it is similar to the overall global solution [101]:

$$V_p = CosineSimilarity\left(X_p, X_g\right) = \frac{X_p.X_g}{\| X_g \| . \| X_p \|} = \frac{\sum_{i=0}^{n-1} X_{p,i} X_{g,i}}{\sqrt{\sum_{i=0}^{n-1} X_{p,i}{}^2} \sqrt{\sum_{i=0}^{n-1} X_{g,i}{}^2}} \tag{5}$$

$$X(t)_{(i,p)}[i]=\begin{cases} X(t-1)_p[i], if\left(S\left(V_i(t)\right)>r\right) \\ X(t-1)_g[i], otherwise \end{cases}, \tag{6}$$

where r is a constant random number in this case. According to Eq (6), the probability of the solution updating its position in the direction of the global solution is higher if it is not a neighbor of the global solution when compared to if it is.

3.2.2. Particle Swarm Optimization (PSO)

BPSO, or binary PSO [106], is based on the fighting strategies used by flocks of birds. Each particle follows the leader particle (global best) and the nearby particles (local best). The particle best solution (pb) refers to the particle's own optimal position. The global best solution (gb) refers to the solution that best fits the swarm as a whole. d is the dimension of the particle. The values of the variables c_1 and c_2 are both set to 1. r_1 *and* r_2 represents a random number between 0 and 1. The number of particles and the number of iterations are both set to 50. The position (Eq (8)) and velocity (Eq (7)) in case of PSO is calculated as follows:

$$v_i^d(t+1)=v_i^d(t)+c_1*r_1*\left(pb_i^d(t)-X_i^d(t)\right)+c_2*r_2*\left(gb^d(t)-X_i^d(t)\right) \tag{7}$$

$$X_i^d(t+1)=X_i^d(t)+v_i^d(t+1) \tag{8}$$

Equation (7) is replaced by Eq (9) for BPSO so that $X_i^d(t)$ can only be either 0 or 1. Here, the sigmoidal function, Eq (10), is used. For BPSO, the position in the binary search space is converted using a sigmoidal function (using Eq (10)), and Eq (8) is replaced by Eq (11). In the case of the BPSO algorithm, the position and velocity of the ith particle are calculated by the following:

$$v_i^d(t+1)=v_i^d(t)*w+c_1*r_1*\left(pb_i^d(t)-X_i^d(t)\right)+c_2*r_2*\left(gb^d(t)-X_i^d(t)\right) \tag{9}$$

$$Sig\left(v_i^d\right)=\frac{1}{1+e^{-v_i^d}} \tag{10}$$

$$X_i^d=\begin{cases} 1, if\ \ Sig\left(v_i^d\right)>r\ and \\ 0, otherwise \end{cases} \tag{11}$$

Additionally, the inertial weight w has a value of 1. *rand* is a random number selected in the interim [0, 1]. $Sig(v_i^d)$ denotes a sigmoidal function, and X_i^d represents the position of the ith particle in dimension d.

3.3. Classification approaches

3.3.1. J48

To create the DT, a modified version of the c4.5 and ID3 algorithms, called J48, is employed [49]. The estimate criteria are used for each node of the DT to choose pertinent input variables for prediction.

The estimate criteria are based on IG and entropy reduction to determine input variables [107]. Equation (12), where *pP* and *pN* represent the fraction of positive and negative (training) instances, yields the entropy (E).

$$E = -pP * log_2(pP) - pN * log_2(pN). \tag{12}$$

There are many different DT algorithms, though one of the most often used ones is probably J48, which is an improved version of the C4.5 tree method and constructs a DT by employing the idea of information entropy [108]. J48 has been extensively used in earlier network security efforts [24,108,109] since it is an integrable component for the machine learning-based security architecture. Therefore, it is a supervised learning model.

3.3.2. Random Forest

The RF method is based on ensemble learning. In the paradigm of "ensemble learning", a learning algorithm may be used repeatedly to improve upon itself.

Since a RF is created by repeatedly running the DT algorithm, it is important to fully comprehend the DT technique before attempting to create one [110]. When cloud network data from smart cities are provided as the input, the DT algorithm's job in the suggested technique is to forecast if class labels are either normal or anomalous. Each tree in the "forest" is created by resampling using the bootstrap methodology. Additionally, a subset of attributes is randomly chosen on each node split, and this subset is used to pick the split variable. For classification, the projected value is the decision of the majority vote. Breiman [50] developed the strategy, which was based on the principles presented by Amit and Geman [111].

One of the most effective techniques used in ML for classification issues is RF. The supervised classification category includes the RF technique. Rather than relying on the result of a single DT, learning is carried out based on the outcomes of many DT [110].

3.3.3. XGBoost

XGBoost is a more recent tree classifier that can scale to large-scale data [112] and is gaining popularity for its outstanding performance across a variety of applications, including cybersecurity (e.g., [24,51,113]). In a nutshell, the classification and regression tree (CART) results are accumulated by Gradient Boosting DT (GBDT) to reach the conclusion. At each iteration, the GBDT must repeatedly traverse the full data collection. The size of the data can only be as much as what can fit in the memory; otherwise, time-consuming read-and-write operations must be performed repeatedly. Therefore, GBDT is unable to satisfy its needs when presented with huge and high-dimensional data. XGBoost was created to address GBDT's problem in handling big samples and high-dimensional data. Tianqi Chen et al. [51] advocated for the creation of XGBoost. In order to achieve high efficiency, versatility, and portability, it is an improved distributed gradient improvement library that applies ML methods within the gradient boosting framework.

Decision trees are generated sequentially by the XGBoost system, an efficient gradient tree-boosting method [51]. It can somewhat perform pertinent calculations in all computer environments more quickly. Because of its effectiveness in modeling newer features and label classification,

XGBoost is widely employed. With the implementation of the XGBoost method in structured and tabular datasets, the use of the technique has greatly increased. The DT-based technique, which involved computing graphical representations of potential decision answers based on specific conditions, served as the foundation for the growth of the XGBoost algorithm. Then, "bagging", which is an ensemble Meta algorithm that aggregates forecasts from several DT using the majoritarian voting technique, was developed. This bagging strategy was further developed to create a forest, or an accumulation of DT, by randomly choosing attributes. The models' performance was improved by lowering the errors that occurred throughout the sequential model generation process. The gradient descent approach was used as an additional improvement to lower the mistakes in the sequential model. Finally, it was determined that the XGBoost algorithm was a useful method for improving the gradient boosting algorithm by removing missing data and eradicating overfitting problems through parallel processing. By utilizing parallelization, tree pruning, and hardware optimization, the XGBoost method optimizes the system.

3.4. WEKA and sark famework

WEKA [48] is a well-known and comprehensive workbench for data mining with an easy-to-use interface. Only a sequential single-node execution is supported. As a result, the size of the datasets and processing jobs that WEKA can manage in its current context is constrained by both sequential execution and the quantity of RAM in a single node. The DistributedWekaSpark may be utilized to circumvent this. It serves as WEKA's distributed framework and preserves the latter's current user interface. The framework is built on top of Spark, a distributed framework linked to Hadoop with quick in-memory processing and support for iterative calculations. WEKA's usability and Spark's processing power are combined to create DistributedWekaSpark, a useable prototype distributed big data mining workbench that executes a variety of real-world scale tasks with an average weak scaling efficiency of 91.4% and an average speed up to 4x quicker than Hadoop [114].

The processing engine Apache Spark is incredibly reliable and scalable. It makes use of a resilient distributed dataset (RDD) [115], which is a group of fault-tolerant components that may be used concurrently. When processing huge datasets in memory, Apache Spark is noted for being quicker than Apache Hadoop MapReduce. Hadoop processes data from the disc, making it ineffective for applications that frequently use repetition in data mining. A more contemporary distributed framework called Spark [40] integrates with Hadoop and offers in-memory computation, which speeds up the processing of iterative jobs, making it a better foundation for data mining. By extending the current WEKA framework, DistributedWekaSpark eliminates the need to completely re-implement algorithms. As a result, existing systems may be more quickly ported, and users can continue to utilize the same interface for both local and remote data processing. In a MapReduce paradigm, it explains a unified framework for representing WEKA's algorithms. As a result, there is no need to examine algorithms to find their parallel components and reimplement them using MapReduce [114]. WEKA developer Mark Hall suggested a trio of additional packages that would give WEKA the ability to perform distributed processing. The first new package, DistributedWekaBase, independently performs fundamental map-reduce functions of any other distributed processing platform. The second one is DistributedWekaHadoop, which offers tasks and wrappers based on the Hadoop platform. The third one is DistributedWekaSpark, which performs tasks based on the Spark platform [116]. The DistributedWekaSpark includes the Spark core classes

that are required and sufficient for local Spark execution on a workstation, communicating with the station's local file system, without the need for a cluster Spark. Additionally, it is possible to run many workers in independent worker threads, taking advantage of all the processors on the computer to maximize power from the project [116].

4. Implementation

As shown in Figure 5, we suggest an architecture that demonstrates the links between IoT-enabled homes and departments, Edge/Fog, and the Cloud, as well as the deployment of IDS at network gateways.

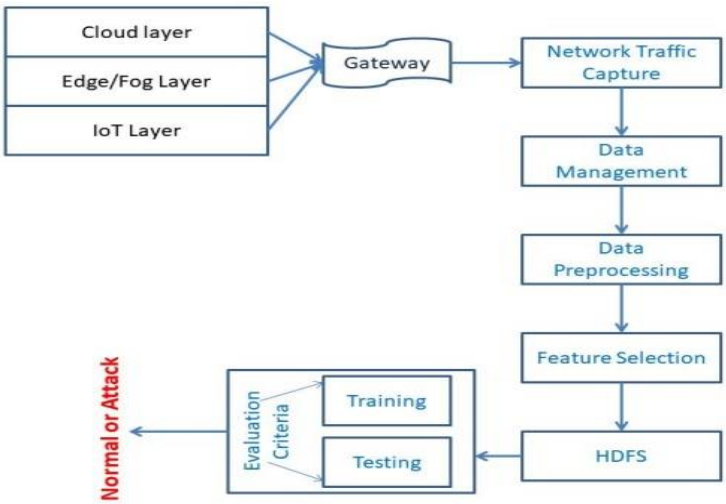


Figure 5. Proposed Methodology with Cloud, Edge/Fog, and IoT layer.

The Edge/Fog and Cloud layers interact with devices and sensors in homes and departments to subscribe to and broadcast telemetry data over network systems. There are several sensors in smart homes and offices, including sensors for the garage, door, smart light, temperature, humidity, and pressure. In Edge/Fog networks, the suggested IDS system would also be installed at gateways, such as routers and switches. It can be used to defend against zero-day attacks on these networks.

The distributed IDS system that is being proposed is used to keep an eye on the endpoints that connect the Edge/Fog, Cloud, and layers of IoT of residences and departments in a smart city, as depicted in Figure 6. The system gathers key network characteristics from these endpoints, logs them in the HDFS, and then adapts its methodology to train and test either normal or attack network vectors.

This section delves into the methods for detecting intrusions. DistributedWekaSpark is used to evaluate the dataset, which is stored in an HDFS. Following the FS technique, we built models using three different classifiers as base classifiers, and one meta-classifier, as shown in Figure 7. These stages are outlined in the sections below.

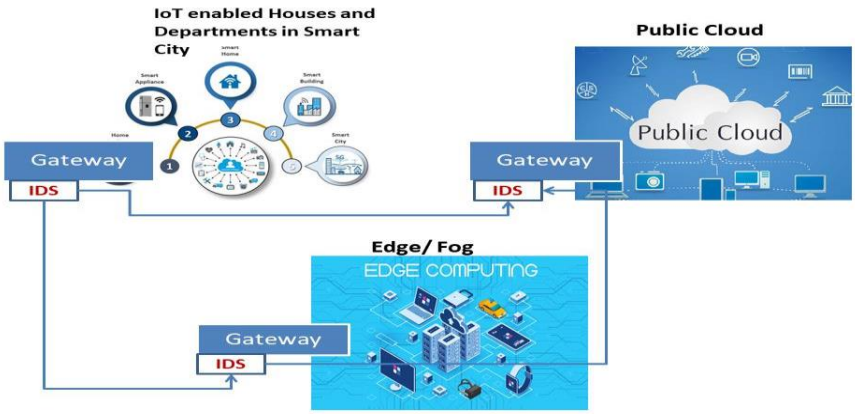


Figure 6. IDS placement in a smart city scenario.

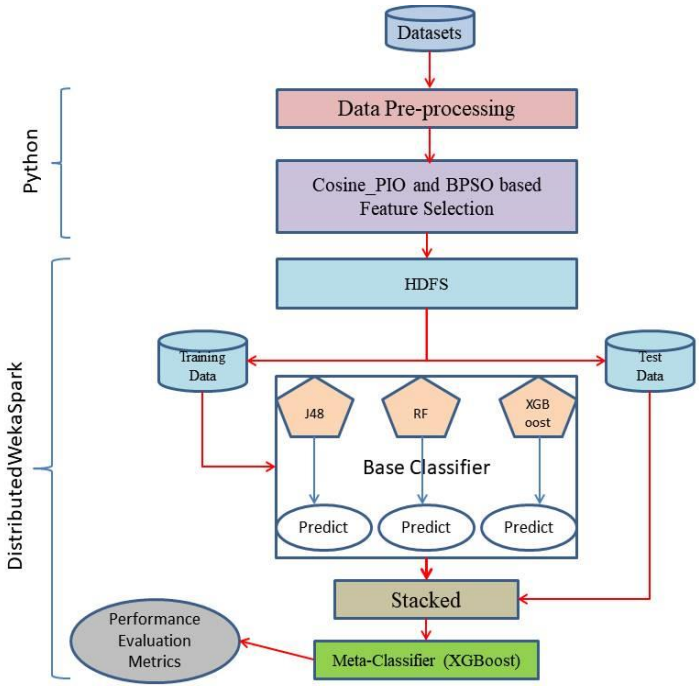


Figure 7. Proposed methods.

4.1. Data preprocessing

Data preprocessing is an important step that may speed up the experiment and enhance the output. Feature normalization and encoding depending on the intrusion dataset's features are part of data pre-processing.

4.1.1. Feature normalization

The range of features is normalized by feature scaling, which guarantees that distinct features have different values. Furthermore, training high-dimensional datasets require high computational

power. Data is frequently scaled using methods such as Z-score standardization, decimal scaling, Max normalization, and Min–Max scaling to address these difficulties [117]. The approach to utilize is frequently determined by the application. Moreover, we have incorporated Min–Max scaling (Eq 13):

$$Min-Maxscalingoffeature X : X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}, \tag{13}$$

where X_{min} and X_{max} are the minimum and maximum values of feature X, respectively.

4.1.2. Feature encoding

For efficient model training, all categorical features will be encoded into vectors. There are several methods for converting categorical data into vectors. 'Label encoding', 'One Hot Encoding', and 'scikit-learn feature mapping' are the most utilized approaches. We adopted the first approach since the number of feature dimensions in the later techniques significantly rises [118]. It took a straightforward approach to convert feature values to numeric numbers; for example, the values of instances like "icmp, http, tcp" in the dataset will turn into vectors 0,1,2, respectively.

4.1.3. Feature removal

Two features (attack_cat and label) are the class labels out of the 45 features in the UNSW-NB15 dataset. The last feature in the NSL-KDD dataset is the class label. Since the objective is to reduce the number of features, it is imperative to get rid of them.

4.2. Feature selection

An adopted metaheuristic based on CPIO and BPSO is used in this article to handle the FS process. In this section, CPIO and BPSO FS techniques were assessed using the NSL-KDD and UNSW-NB15 datasets. The CPIO approach's chosen collection of features from the NSL-KDD and UNSW-NB15 datasets is shown in Table 6. All the aforementioned FS approaches are carried out using Python, on a workstation using a 64-bit Windows operating system and a 2.40 GHz Intel Xeon processor and 16 GB of RAM.

Table 6. Selected features.

Dataset	Approach	No. of features Selected	Feature Number
NSL-KDD	CPIO	5	27, 22, 10, 6, 2
	BPSO	9	39, 37, 30, 29, 26, 12, 6, 5, 4
UNSW-NB15	CPIO	5	29, 12, 8, 4, 3
	BPSO	16	43, 33, 29, 27, 26, 25, 24, 21, 17, 16, 12, 11, 7, 5, 2, 1

We investigate only five and nine of the 41 features in the NSL-KDD dataset based on CPIO and BPSO, respectively, and only five and 16 of the 43 features in the UNSW-NB15 dataset based on CPIO and BPSO, respectively. By reducing the number of features, the smaller subset of features

may assist us in designing a simpler model. Additionally, the model's detection skills are improved by removing redundant features. Once the FS procedure has been completed using the FS algorithms, the collection of features is trained using an SEM for classification.

4.3. Stacking Ensemble Method (SEM)

Ensemble approaches are a type of ML methodology in which numerous base classifiers are combined to generate a single, effective prediction model [43,119]. The final model will overcome each learner's flaws, yielding a strong model that will improve prediction results. Algorithm 1 explains the procedures necessary for training our proposed SEM.

The SEM is a general architecture made up of two types of classifiers: base and meta-classifiers. The training dataset is used to train the base (initial) classifiers, while a new dataset is created for the meta-classifier. Then, this new dataset is used to train the meta-classifier. Finally, the test dataset is predicted using the trained meta-classifier. We provide a model based on the SEM of ML algorithms, in which J48, RF, and XGBoost serve as base classifiers, and XGBoost serves as a meta-classifier. This research supports all of the proposed classifiers, particularly because their findings are simply interpretable, and their training is robust against outliers.

Algorithm 1 SEM

Input: Training Data $T\{X_i, Y_i\}_{i=1}^{a}$ where $X = X_i \epsilon S^b$ is a give record set and $Y = Y_i \epsilon N$ is a label set.
Output: $Ensemble\ E's\ prediction$

Begin
Step 1: Divide *T* into '*a*' equal size subset randomly, i.e., T = {T_1, T_2, T_3,……..T_a}.
Step 2:
for *a* ←1 to *A*
Learn base classifiers namely, J48, RF and XGBoost
for *b* ←1 to *B*
Learn a base classifier F_{ab} from T or T_a
end for
Step 3: Generate a meta-classifier (XGBoost) training dataset
for each $X_i \epsilon T_a$
Extract a new instance (x'_i, y_i) where $x'_i=\{F_{a1}(X_i), F_{a2}(X_i), F_{a3}(X_i),\ldots, F_{aB}(X_i)\}$
end for
end for
Return $y_i = \{y_1, y_2, \ldots\ldots\ldots, y_b\}$ for ensemble.
End

4.4. Performance assessment

The most popular performance metrics, including sensitivity, specificity, precision, FPR, accuracy, F1 Score, and MCC, were utilized to assess the performance. Table 7 represents the confusion matrix, which displays how well a classification system performs. The undermentioned metrics in Table 8 are widely used to assess models. The following are the performance metrics. True Positive (TP) refers to an attack sample that has been correctly identified as an attack. A specimen that is correctly identified as normal is represented by the True Negative (TN) code. False Positive

(FP) refers to the misidentification of an attack in a normal specimen. An attack sample that has been incorrectly classified as normal is known as a False Negative (FN).

Table 7. Confusion matrix.

		Actual	
		Benign	Malware
Predicted	Benign	TP	FP
	Malware	FN	TN

Table 8. Metrics generated from the confusion matrix for performance evaluation.

Metrics	Formula
MCC	"(TP*TN-FP*FN)/sqrt((TP+FP)*(TP+FN)*(TN+FP)*(TN+FN))"
Accuracy	"(TP+TN)/(FP+TP+FN+TN)"
Precision	"TP/(TP+FP)"
Sensitivity	"TP/(TP+FN)"
F1-Score	"2TP/(2TP+FP+FN)"
FPR	"FP/(FP+TN)"
Specificity	"TN/(FP+TN)"

4.5. Model performance

Using the features chosen by CPIO and BPSO for the NSL-KDD and UNSW-NB15 datasets, Figures 8 and 9 and Tables 9−12 show the outcomes of the proposed methodology by distinguishing between classes that are either attack or normal for the supplied dataset. By training the model with only the chosen features, each FS strategy or method was tested using the base classifiers and SEM. Then, the model was examined using the testing set. An average of 20 runs was used to calculate the results. However, our analysis considered a more trustworthy metric (such as MCC) that was discovered to produce more accurate estimates for the suggested model. Therefore, our study argues for the use of the MCC metric as an evaluation criterion in future work, particularly in anomaly-based IDS.

In the case of NSL-KDD, the results of stacking-based classification appear promising for identifying intrusions in the Cloud network data coming from smart cities when using the feature chosen by CPIO and BPSO. The best classification result for the features selected by CPIO, sensitivity (0.9730), specificity and precision (0.9852), accuracy (0.9791), F1-Score (0.9790), MCC (0.9582), and FPR (0.0148), is for SEM, as depicted in Table 9 and Figure 8. The best classification result for the features selected by BPSO, sensitivity (0.9810), specificity and precision (0.9923), accuracy (0.9862), F1-Score (0.9861), MCC (0.9724), and FPR (0.0077), is for SEM, as depicted in Table 10 and Figure 8.

Table 9. Finding for NSL-KDD dataset using CPIO selected features.

Classifier	Sensitivity	Specificity	Precision	FPR	Accuracy	F1-Score	MCC
J48	0.9289	0.9406	0.9405	0.0594	0.9347	0.9346	0.8695
RF	0.9377	0.9495	0.9494	0.0505	0.9436	0.9435	0.8872
XGBoost	0.9553	0.9674	0.9673	0.0326	0.9613	0.9613	0.9227
Stacking	0.9730	0.9852	0.9852	0.0148	0.9791	0.9790	0.9582

Table 10. Finding for NSL-KDD dataset using BPSO selected features.

Classifier	Sensitivity	Specificity	Precision	FPR	Accuracy	F1-Score	MCC
J48	0.9359	0.9477	0.9476	0.0523	0.9418	0.9417	0.8837
RF	0.9474	0.9593	0.9592	0.0407	0.9533	0.9533	0.9067
XGBoost	0.9580	0.9700	0.9700	0.0300	0.9640	0.9639	0.9280
Stacking	0.9801	0.9923	0.9923	0.0077	0.9862	0.9861	0.9724

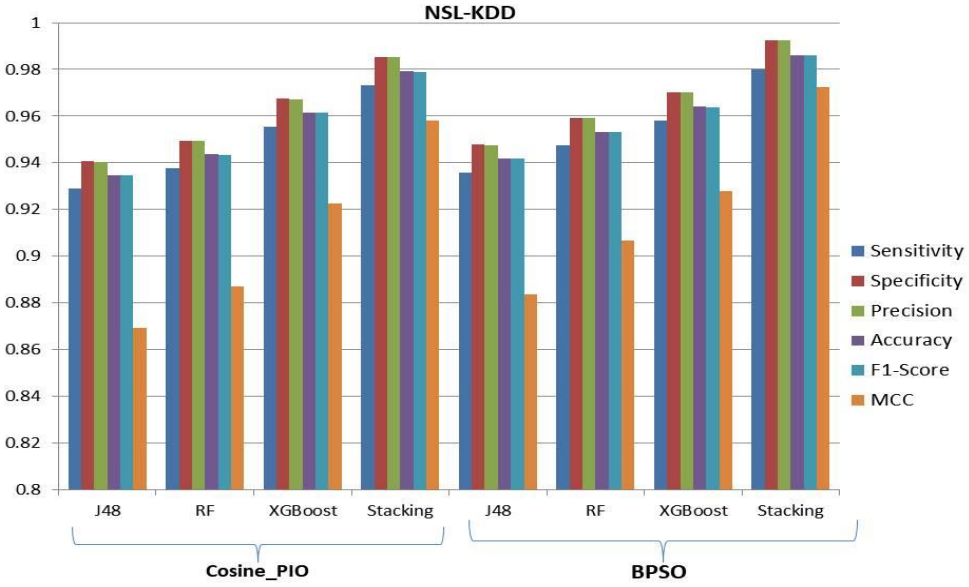


Figure 8. Results for both FS methods on the NSL-KDD dataset.

For NSL-KDD, the results of SEM-based classification appear promising for identifying intrusion. A difference of about 0.8188% in sensitivity when the feature selected by BPSO is considered. BPSO selected features generate better classification results, since there is a difference of 0.7225% in accuracy. When CPIO selected features are considered for classification, a higher FPR is obtained; when compared to the consideration of features selected by BPSO, the percentage difference is 63.11% in the FPR.

In the case of UNSW-NB15, the results of stacking-based classification appear promising for identifying intrusions in the cloud network data coming from smart cities when using the feature chosen by CPIO and BPSO. The best classification result for the features selected by CPIO, sensitivity (0.9562), specificity (0.9553), precision (0.9552), accuracy (0.9558), F1-Score (0.9557), MCC (0.9115), and FPR (0.0337), is for SEM, as depicted in Table 11 and Figure 9. The best classification result for the features selected by BPSO, sensitivity (0.9587), specificity (0.9577), precision (0.9577), accuracy (0.9582), F1-Score (0.9582), MCC (0.9164), and FPR (0.0323), is for SEM, as depicted in Table 12 and Figure 9.

For UNSW-NB15, a difference of about 0.2611% in sensitivity when the feature selected by BPSO is considered. BPSO selected features generates better classification results as there is a difference of 0.2507% in accuracy. When CPIO selected features are considered for classification, a higher FPR is received; when compared to the consideration of features selected by BPSO, the difference is 4.24% in the FPR.

Table 11. Finding for UNSW-NB15 dataset using CPIO selected features.

Classifier	Sensitivity	Specificity	Precision	FPR	Accuracy	F1-Score	MCC
J48	0.9404	0.9395	0.9394	0.0605	0.9400	0.9399	0.8799
RF	0.9441	0.9431	0.9431	0.0569	0.9436	0.9436	0.8872
XGBoost	0.9489	0.9480	0.9479	0.0520	0.9485	0.9484	0.8969
Stacking	0.9562	0.9553	0.9552	0.0337	0.9558	0.9557	0.9115

Table 12. Finding for UNSW-NB15 dataset using BPSO selected features.

Classifier	Sensitivity	Specificity	Precision	FPR	Accuracy	F1-Score	MCC
J48	0.9429	0.9419	0.9419	0.0581	0.9424	0.9424	0.8848
RF	0.9477	0.9468	0.9467	0.0532	0.9473	0.9472	0.8945
XGBoost	0.9526	0.9516	0.9516	0.0484	0.9521	0.9521	0.9042
Stacking	0.9587	0.9577	0.9577	0.0323	0.9582	0.9582	0.9164

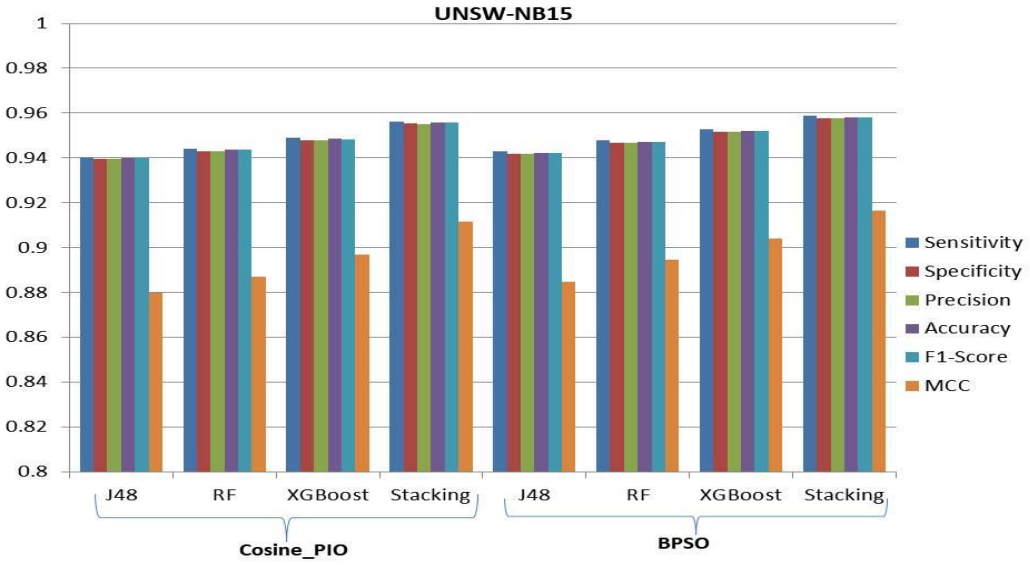


Figure 9. Results for both FS methods on the UNSW-NB15 dataset.

The stacking approach requires more processing time, since it combines several base classifiers, each of which requires development time. The amount of time it takes classifiers to forecast intrusions for the test dataset is shown in Table 13. In terms of model building and testing time, we found that the best classifiers in our setting are J48 and RF, and J48 obtains the lowest computation time even if the complexity of the stacking model has risen; as a result, the time requirements increased, which beats conventional IDS, as noted in the previous result, and is a significant consideration. If computationally expensive, high-performing classification strategies have significant implications for IoT-based smart cities applications. The cost of missing an intrusion in such a system can be quite expensive. As an outcome, the cost of a little extra time, which is reported in seconds for the datasets examined and therefore potentially well scalable in comparison to earlier approaches, is justified. The capacity to quickly identify odd activity in the network is crucial for the sustainability of services in commercial sectors such as smart cities and financial institutions. Attacks

that go unnoticed in these places can be expensive, though manually identifying the attacks can be exceedingly challenging. The focus is on a precise intrusion detection in such systems, which frequently use considerable computational resources for automatically identifying intrusions. As a result, the suggested model has significant practical usefulness. Table 13 shows the model construction time for the given dataset. J48 takes the least amount of time among the classifiers, whereas stacking takes the most time to create models for both datasets.

Table 13. for the base and stacking ensembles, model construction, and testing time.

Methodologies	NSL-KDD		UNSW-NB15	
	Model Building Time (s)	Testing Time (μs)	Model Building Time (s)	Testing Time (μs)
J48	0.458	0.187	0.97	0.436
RF	0.593	0.287	1.30	0.487
XGBoost	1.21	0.79	2.01	1.14
Stacking	6.34	3.09	10.54	5.74

4.6. Performance comparison with current methodologies

Using the NSL-KDD and UNSW-NB15 datasets, Table 14 compares the performance of the stacking model with other methodologies. The proposed model outperforms earlier similar ensemble classifiers described in [97], which, according to the table, use 10-fold cross validation and consider intrusion detection as a classification issue. In terms of accuracy, our spark-based SEM-oriented ensemble model exceeds several current approaches. A "-" in the table denotes a value that is either inapplicable or unavailable.

For comparing the outcomes, we primarily employ the accuracy and FPR variables. This is a tried-and-true method that has been applied to a variety of practical machine-learning projects. We have compared the achieved rates with those presented in undermentioned work because accuracy rates are a crucial component of any IDS performance evaluation. Since our method uses a nature-inspired FS approach and classifiers that are HDFS and Spark-based, which is not the case with the other models, it outperforms the other models' accuracy rates. FPR is a term used to explain the inability to recognize normal behavior. In other words, there is a warning. The table below compares the FPR of our methodology to the works described in the citations. Compared to the current state-of-the-art, our method yields the lowest FPR: 0.0077% for NSL-KDD and 0.0323% for UNSW-NB15.

Our approach is distributed in nature with the aid of HDFS and DistributedWekaSpark, thereby ensuring high availability and fault tolerance of our IDS and making it appropriate to handle big Cloud network data of smart cities, which is another crucial point that sets it apart from all the works cited. Our presented methodology outperforms [101] in terms of accuracy and the number of selected features, while using the same number of features. Additionally, the comparative results demonstrate that our work surpasses many other works in terms of accuracy and FPR when using the NSL-KDD and UNSW-NB15 datasets, as shown in Table 14. In comparison to [95], we can see that our research is able to identify fewer features for the NSL-KDD and UNSW-NB15 datasets, with the latter displays a superior accuracy. Summarizing the results, we can see that our approach outperformed all other methods in terms of accuracy. The suggested model makes it abundantly

evident that the presented approach outperforms earlier reported approaches in terms of results.

Table 14. Performance comparison with current methodologies.

NSL-KDD						
Author & Reference	Year	Methodologies	FS Approach	No. of Features Selected	Accuracy	FPR
Alazzam et al. [101]	2021	DT	SPIO	18	0.869	0.064
			CPIO	5	0.883	0.088
Khraisat et al. [87]	2020	C5-DT/OC-SVM	-	-	0.8324	-
Tama et al. [95]	2019	REPT	PSO	37	0.8579	11.7
Louk and Tama [88]	2023	Bagging-GBM	-	-	0.9157	1.3
Krishnaveli et al. [89]	2022	Weighted majority voting	-	-	0.8523	12.8
Zhang et al. [90]	2021	MFFSEM	-	-	0.8433	24.82
Tama et al. [91]	2020	Stacking	-	-	0.9217	2.52
Prabavathy et al. [76]	2018	OS-ELM	-	-	0.9736	0.37
Shrivas et al. [81]	2014	ANN-Bayesian	GR	29	0.9778	-
Zhou et al. [47]	2020	Voting Ensemble	CFS-BA	10	0.8737	3.19
Salo et al. [97]	2019	Ensemble	IG-PCA	12	0.9824	0.017
Alghanam et al. [92]	2021	LS-PIO	iForest	10	0.947	-
Proposed Work	-	SEM (HDFS and DistributedWekaSpark)	CPIO	5	0.9791	0.0148
			BPSO	9	0.9862	0.0077
UNSW-NB15						
Alazzam et al. [101]	2021	DT	SPIO	14	0.913	0.052
			CPIO	5	0.917	0.034
Rashid et al. [45]	2022	Ensemble	SelectKbest	20	0.94	0.06
Smitha et al. [96]	2020	Stacking Ensemble	-	42	0.9400	5.2
Tama et al. [95]	2019	REPT	PSO	19	0.9127	8.90
Alghanam et al. [92]	2021	LS-PIO	iForest	10	0.9445	-
Zehong et al. [93]	2022	EFS-DNN	Light-GBM	15	0.8834	12.46
Nazir et al. [95]	2021	TS-RF	TS	16	0.8312	3.7
Proposed Work	-	SEM (HDFS and DistributedWekaSpark)	CPIO	5	0.9558	0.0337
			BPSO	16	0.9582	0.0323

5. Conclusions

In this study, we developed a distributed and potent IDS that enables the processing of large amounts of Cloud data from Smart Cities and improves accuracy while utilizing the fewest features possible. It uses Spark and ML approaches to effectively manage massive amounts of data in vast networks of smart cities. We used the Python-based FS methods CPIO and BPSO to create this

system. The IDS used in this study for Cloud network data from smart cities used Spark and WEKA. Due to the connection between WEKA and Spark (DistributedWekaSpark package), it is distributed and scalable. Using the capabilities of distributed systems while maintaining the familiar WEKA interface, DistributedWekaSpark is a scalable Big Data Mining toolkit. Built on top of Spark, DistributedWekaSpark offers quick in-memory iterative processing using both parallel and distributed execution, making it the perfect platform for data mining techniques. Using WEKA's Knowledge flow, this combination enables the analysis of Cloud network data for smart cities and the storage of HDFS data. In order to build parallelized learning models for cyber-data analytics, we used machine-learning approaches for feature extraction and selection. For NSL-KKD and UNSW-NB15, the CPIO FS technique reduced the number of selected features from 41 to five and from 43 to five features, respectively. For NSL-KKD and UNSW-NB15, the BPSO FS technique reduced the number of selected features from 41 to nine and from 43 to 16 features, respectively. For classifying the cloud network traffic of smart cities, the tree-based SEM of J48, RF, and XGBoost was applied. The best results were obtained for sensitivity (0.9810), specificity and precision (0.9923), accuracy (0.9862), F1-Score (0.9861), MCC (0.9724), and FPR (0.0077) in the NSL-KDD dataset, while in case of UNSW-NB15 dataset, the best results were obtained for sensitivity (0.9587), specificity (0.9577), precision (0.9577), accuracy (0.9582), F1-Score (0.9582), MCC (0.9164), and FPR (0.0323). The results demonstrate that CPIO and BPSO contribute to a greater accuracy and better outcomes fitting. Since Spark functionality has been implemented, our methodology has been discovered to be scalable and dispersed, making it suitable for the IoT context of smart cities. Compared to contemporary systems, our suggested system experimentally exhibits a higher accuracy and lower FPR.

As a result, research in the future must take a larger range of intrusion data sets in diverse settings, environments and with wider range of threats into account. More evaluation metrics will be used in upcoming studies. Using upgraded and latest nature-inspired algorithms for FS and several deep neural network algorithms, including Auto-Encoder, Gated Recurrent Units and LSTM, will be used to implement the strategy. We intend to use explainable AI for IDS ML/DL-based algorithms for the detection and classification of cyberattacks in networks of smart cities.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgements

The first author would like to thank the Department of Science & Technology (DST), Ministry of Science and Technology, GOI, for financial support in terms of research fellowship (DST-INSPIRE).

Conflict of interest

The authors declare that they do not have any known competing interests.

References

1. Z. Ullah, F. Al-Turjman, L. Mostarda, R. Gagliardi, Applications of artificial intelligence and machine learning in smart cities, *Comput. Commun.*, **154** (2020), 313–323. https://doi.org/10.1016/j.comcom.2020.02.069
2. Urbanization, 2023. Available from: https://www.unfpa.org/urbanization.
3. R. Petrolo, V. Loscrì, N. Mitton, Towards a smart city based on cloud of things, a survey on the smart city vision and paradigms, *Trans. Emerg. Telecommun. Technol.*, **28** (2017). https://doi.org/10.1002/ETT.2931
4. U. Aguilera, O. Peña, O. Belmonte, D. López-de-Ipiña, Citizen-centric data services for smarter cities, *Future Gener. Comput. Syst.*, **76** (2017), 234–247. https://doi.org/10.1016/j.future.2016.10.031
5. P. Neirotti, A. De Marco, A. C. Cagliano, G. Mangano, F. Scorrano, Current trends in smart city initiatives: some stylised facts, *Cities*, **38** (2014), 25–36. https://doi.org/10.1016/j.cities.2013.12.010
6. H. Habibzadeh, B. H. Nussbaum, F. Anjomshoa, B. Kantarci, T. Soyata, A survey on cybersecurity, data privacy, and policy issues in cyber-physical system deployments in smart cities, *Sustain. Cities Soc.*, **50** (2019), 101660. https://doi.org/10.1016/J.SCS.2019.101660
7. M. Pouryazdan, C. Fiandrino, B. Kantarci, T. Soyata, D. Kliazovich, P. Bouvry, Intelligent gaming for mobile crowd-sensing participants to acquire trustworthy big data in the Internet of Things, *IEEE Access*, **5** (2017), 22209–22223. https://doi.org/10.1109/ACCESS.2017.2762238
8. K. Liao, Z. Zhao, A. Doupe, G. J. Ahn, Behind closed doors: measurement and analysis of CryptoLocker ransoms in Bitcoin, in *2016 APWG Symposium on Electronic Crime Research (eCrime)*, **2016** (2016), 1–13. https://doi.org/10.1109/ECRIME.2016.7487938
9. K. Cabaj, W. Mazurczyk, Using software-defined networking for ransomware mitigation: the case of cryptowall, *IEEE Netw.*, **30** (2016), 14–20. https://doi.org/10.1109/MNET.2016.1600110NM
10. C. Miller, C. Valasek, Remote exploitation of an unaltered passenger vehicle, Black Hat USA, 2015. Available from: https://ioactive.com/wp-content/uploads/2018/05/IOActive_Remote_Car_Hacking-1.pdf.
11. A. Greenberg, Hackers remotely kill a jeep on the highway—with me in it, *Wired*, **7** (2015), 21–22.
12. N. Moustafa, M. Keshk, K. K. R. Choo, T. Lynar, S. Camtepe, M. Whitty, DAD: a distributed anomaly detection system using ensemble one-class statistical learning in edge networks, *Future Gener. Comput. Syst.*, **118** (2021), 240–251. https://doi.org/10.1016/J.FUTURE.2021.01.011
13. T. Alam, Cloud-based IoT applications and their roles in smart cities, *Smart Cities*, **4** (2021), 1196–1219. https://doi.org/10.3390/smartcities4030064
14. Y. Liu, C. Yang, L. Jiang, S. Xie, Y. Zhang, Intelligent edge computing for IoT-based energy management in smart cities, *IEEE Netw.*, **33** (2019), 111–117. https://doi.org/10.1109/MNET.2019.1800254.
15. Z. Allam, Z. A. Dhunny, On big data, artificial intelligence and smart cities, *Cities*, **89** (2019), 80–91. https://doi.org/10.1016/j.cities.2019.01.032
16. H. Habibzadeh, T. Soyata, B. Kantarci, A. Boukerche, C. Kaptan, Sensing, communication and security planes: a new challenge for a smart city system design, *Comput. Netw.*, **144** (2018), 163–200. https://doi.org/10.1016/J.COMNET.2018.08.001

17. M. A. Rahman, A. T. Asyhari, L. S. Leong, G. B. Satrya, M. H. Tao, M. F. Zolkipli, Scalable machine learning-based intrusion detection system for IoT-enabled smart cities, *Sustain. Cities Soc.*, **61** (2020), 102324. https://doi.org/10.1016/J.SCS.2020.102324
18. H. H. Pajouh, R. Javidan, R. Khayami, A. Dehghantanha, K. K. R. Choo, A two-layer dimension reduction and two-tier classification model for anomaly-based intrusion detection in IoT backbone networks, *IEEE Trans. Emerging Top. Comput.*, **7** (2019), 314–323. https://doi.org/10.1109/TETC.2016.2633228
19. M. E. Aminanto, R. Choi, H. C. Tanuwidjaja, P. Yoo, K. Kim, Deep abstraction and weighted feature selection for Wi-Fi impersonation detection, *IEEE Trans. Inf. Forensics Secur.*, **13** (2017), 621–636. https://doi.org/10.1109/TIFS.2017.2762828
20. C. F. Tsai, Y. F. Hsu, C. Y. Lin, W. Y. Lin, Intrusion detection by machine learning: a review, *Expert Syst. Appl.*, **36** (2009), 11994–12000. https://doi.org/10.1016/j.eswa.2009.05.029
21. A. L. Buczak, E. Guven, A survey of data mining and machine learning methods for cyber security intrusion detection, *IEEE Commun. Surv. Tutorials*, **18** (2015), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502
22. Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, et al., Machine learning and deep learning methods for cybersecurity, *IEEE Access*, **6** (2018), 35365–35381. https://doi.org/10.1109/ACCESS.2018.2836950
23. L. Tian, Design and implementation of a distributed intelligent network intrusion detection system, in *2010 Int. Conf. Electr. Control Eng.*, **2010** (2010), 683–686. https://doi.org/10.1109/ICECE.2010.174
24. C. Kolias, G. Kambourakis, A. Stavrou, S. Gritzalis, Intrusion detection in 802.11 networks: empirical evaluation of threats and a public dataset, *IEEE Commun. Surv. Tutorials*, **18** (2016), 184–208. https://doi.org/10.1109/COMST.2015.2402161.
25. A. A. Aryachandra, Y. F. Arif, S. N. Anggis, Intrusion Detection System (IDS) server placement analysis in cloud computing, in *2016 4th Int. Conf. Inform. Commun. Technol. (ICoICT)*, **2016** (2016). https://doi.org/10.1109/ICOICT.2016.7571954
26. D. B. Rawat, K. Z. Ghafoor, *Smart Cities Cybersecurity and Privacy*, Elsevier, 2018.
27. N. Sengupta, Designing cyber security system for smart cities, in *Smart Cities Symposium 2018*, **2018** (2018). https://doi.org/10.1049/cp.2018.1418
28. E. Vasilomanolakis, S. Karuppayah, M. Muhlhauser, M. Fischer, Taxonomy and survey of collaborative intrusion detection, *ACM Comput. Surv.*, **47** (2015), 1−33. https://doi.org/10.1145/2716260
29. H. Liu, H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, Springer Science & Business Media, 2012. https://doi.org/10.1007/978-1-4615-5689-3
30. X. Tang, Y. Dai, Y. Xiang, Feature selection based on feature interactions with application to text categorization, *Expert Syst. Appl.*, **120** (2019), 207–216. https://doi.org/10.1016/j.eswa.2018.11.018
31. S. Mohammadi, H. Mirvaziri, M. Ghazizadeh-Ahsaee, H. Karimipour, Cyber intrusion detection by combined feature selection algorithm, *J. Inf. Secur. Appl.*, **44** (2019), 80–88. https://doi.org/10.1016/j.jisa.2018.11.007
32. S. Maza, M. Touahria, Feature selection algorithms in intrusion detection system: a survey, *KSII Trans. Internet Inf. Syst.*, **12** (2018), 5079–5099. https://doi.org/10.3837/tiis.2018.10.024
33. A. Al Shorman, H. Faris, I. Aljarah, Unsupervised intelligent system based on one class support vector machine and Grey Wolf optimization for IoT botnet detection, *J. Ambient Intell. Hum. Comput.*, **11** (2020), 2809–2825. https://doi.org/10.1007/s12652-019-01387-y

34. H. Alazzam, A. Sharieh, K. E. Sabri, A feature selection algorithm for intrusion detection system based on pigeon inspired optimizer, *Expert Syst. Appl.*, **148** (2020), 113249. https://doi.org/10.1016/J.ESWA.2020.113249
35. R. Poli, J. Kennedy, T. Blackwell, Particle swarm optimization, *Swarm Intell.*, **1** (2007), 33–57. https://doi.org/10.1007/s11721-007-0002-0
36. A. Jain, V. Sharma, V. Sharma, Big data mining using supervised machine learning approaches for Hadoop with Weka distribution, *Int. J. Comput. Intell. Res.*, **13** (2017), 2095–2111.
37. M. R. Ghazi, D. Gangodkar, Hadoop, MapReduce and HDFS: a developers perspective, *Procedia Comput. Sci.*, **48** (2015), 45–50. https://doi.org/10.1016/j.procs.2015.04.108
38. M. R. Ghazi, N. S. Raghava, MapReduce based analysis of sample applications using hadoop, in *Int. Conf. Appl. Comput. Commun. Technol.*, Springer, **899** (2018), 34–44. https://doi.org/10.1007/978-981-13-2035-4_4
39. M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, I. Stoica, Spark: cluster computing with working sets, **10** (2010), 1−7.
40. A. G. Shoro, T. R. Soomro, Big data analysis: apache spark perspective, *Global J. Comput. Sci. Technol.*, **15** (2015), 7–14.
41. A. K. Saxena, S. Sinha, P. Shukla, General study of intrusion detection system and survey of agent based intrusion detection system, in *2017 Int. Conf. Comput., Commun. Automation (ICCCA)*, **2017** (2017), 421–471. https://doi.org/10.1109/CCAA.2017.8229866
42. I. H. Sarker, Y. B. Abushark, F. Alsolami, A. I. Khan, Intrudtree: a machine learning based cyber security intrusion detection model, *Symmetry*, **12** (2020), 754. https://doi.org/10.3390/sym12050754
43. L. K. Hansen, P. Salamon, Neural network ensembles, *IEEE Trans. Pattern Anal. Mach. Intell.*, **12** (1990), 993–1001. https://doi.org/10.1109/34.58871
44. N. T. Pham, E. Foo, S. Suriadi, H. Jeffrey, H. F. M. Lahza, Improving performance of intrusion detection system using ensemble methods and feature selection, in *Proceedings of the Australasian Computer Science Week Multiconference*, **2018** (2018), 1–6. https://doi.org/10.1145/3167918.3167951
45. M. Rashid, J. Kamruzzaman, T. Imam, S. Wibowo, S. Gordon, A tree-based stacking ensemble technique with feature selection for network intrusion detection, *Appl. Intell.*, **52** (2022), 9768–9781. https://doi.org/10.1007/s10489-021-02968-1
46. I. H. Sarker, A. S. M. Kayes, S. Badsha, H. Alqahtani, P. Watters, A. Ng, Cybersecurity data science: an overview from machine learning perspective, *J. Big Data*, **7** (2020), 41. https://doi.org/10.1186/s40537-020-00318-5
47. Y. Zhou, G. Cheng, S. Jiang, M. Dai, Building an efficient intrusion detection system based on feature selection and ensemble classifier, *Comput. Netw.*, **174** (2020), 107247. https://doi.org/10.1016/j.comnet.2020.107247
48. E. Frank, M. A. Hall, I. H. Witten, The WEKA Workbench, Online appendix for "data mining: practical machine learning tools and techniques", Morgan Kaufmann, 2016. Available from: https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf.
49. J. R. Quinlan, *C4.5: Programs for Machine Learning*, Elsevier, 2014. Available from: https://books.google.com/books/about/C4_5.html?id=b3ujBQAAQBAJ.
50. L. Breiman, Random Forests, *Mach. Learn.*, **45** (2001), 5–32. https://doi.org/10.1023/A:1010933404324

51. T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, **2016** (2016), 785–794. https://doi.org/10.1145/2939672.2939785
52. N. Moustafa, J. Slay, UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set), in *2015 Military Communications and Information Systems Conference (MilCIS)*, **2015** (2015). https://doi.org/10.1109/MILCIS.2015.7348942.
53. M. Tavallaee, E. Bagheri, W. Lu, A. A. Ghorbani, A detailed analysis of the KDD CUP 99 data set, in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, **2009** (2009). https://doi.org/10.1109/CISDA.2009.5356528
54. L. P. Qian, Y. Wu, B. Ji, L. Huang, D. H. K. Tsang, HybridIoT: integration of hierarchical multiple access and computation offloading for IoT-based smart cities, *IEEE Netw.*, **33** (2019), 6–13. https://doi.org/10.1109/MNET.2019.1800149
55. S. Garg, A. Singh, S. Batra, N. Kumar, L. T. Yang, UAV-empowered edge computing environment for cyber-threat detection in smart vehicles, *IEEE Netw.*, **32** (2018), 42–51. https://doi.org/10.1109/MNET.2018.1700286
56. M. Dener, The role of cloud computing in smart cities, in *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)*, **7** (2019), 39–43.
57. M. Chen, W. Liu, T. Wang, S. Zhang, A. Liu, A game-based deep reinforcement learning approach for energy-efficient computation in MEC systems, *Knowl.-Based Syst.*, **235** (2022), 107660. https://doi.org/10.1016/j.knosys.2021.107660
58. X. Zhu, Y. Luo, A. Liu, N. N. Xiong, M. Dong, S. Zhang, A deep reinforcement learning-based resource management game in vehicular edge computing, *IEEE Trans. Intell. Transp. Syst.*, **23** (2022), 2422–2433. https://doi.org/10.1109/TITS.2021.3114295
59. H. A. Khattak, H. Farman, B. Jan, I. U. Din, Toward integrating vehicular clouds with IoT for smart city services, *IEEE Netw.*, **33** (2019), 65–71. https://doi.org/10.1109/MNET.2019.1800236
60. M. Kaur, P. Maheshwari, Building smart cities applications using IoT and cloud-based architectures, in *2016 Int. Conf. Ind. Inform. Comput. Syst. (CIICS)*, **2016** (2016), 1–5. https://doi.org/10.1109/ICCSII.2016.7462433
61. R. Massobrio, S. Nesmachnow, A. Tchernykh, A. Avetisyan, G. Radchenko, Towards a cloud computing paradigm for big data analysis in smart cities, *Program. Comput. Software*, **44** (2018), 181–189. https://doi.org/10.1134/S0361768818030052.
62. L. A. B. Pacheco, E. A. P. Alchieri, P. A. S. M. Barreto, Device-based security to improve user privacy in the Internet of Things, *Sensors*, **18** (2018). https://doi.org/10.3390/s18082664
63. S. Chawla, Deep learning based intrusion detection system for Internet of Things, University of Washington, 2017. Available from: https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/39829/Chawla_washington_0250O_17062.pdf.
64. I. Alrashdi, A. Alqazzaz, E. Aloufi, R. Alharthi, M. Zohdy, H. Ming, AD-IoT: anomaly detection of IoT cyberattacks in smart city using machine learning, in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, **2019** (2019), 305–310. https://doi.org/10.1109/CCWC.2019.8666450
65. A. Elsaeidy, I. Elgendi, K. S. Munasinghe, D. Sharma, A. Jamalipour, A smart city cyber security platform for narrowband networks, in *2017 27th Int. Telecommun. Netw. Appl. Conf. (ITNAC)*, **2017** (2017), 1–6. https://doi.org/10.1109/ATNAC.2017.8215388

66. A. A. Alli, M. M. Alam, SecOFF-FCIoT: machine learning based secure offloading in Fog-Cloud of things for smart city applications, *Internet Things*, **7** (2019), 100070. https://doi.org/10.1016/J.IOT.2019.100070
67. M. Aloqaily, S. Otoum, I. Al Ridhawi, Y. Jararweh, An intrusion detection system for connected vehicles in smart cities, *Ad Hoc Netw.*, **90** (2019), 101842. https://doi.org/10.1016/J.ADHOC.2019.02.001
68. H. Sedjelmaci, S. M. Senouci, M. Al-Bahri, A lightweight anomaly detection technique for low-resource IoT devices: a game-theoretic methodology, in *2016 IEEE Int. Conf. Commun. (ICC)*, IEEE, **2016** (2016), 1–6. https://doi.org/10.1109/ICC.2016.7510811
69. D. H. Summerville, K. M. Zach, Y. Chen, Ultra-lightweight deep packet anomaly detection for Internet of Things devices, in *2015 IEEE 34th International Performance Computing and Communications Conference (IPCCC)*, **2015** (2015), 1–8. https://doi.org/10.1109/PCCC.2015.7410342
70. H. Bostani, M. Sheikhan, Hybrid of binary gravitational search algorithm and mutual information for feature selection in intrusion detection systems, *Soft Comput.*, **21** (2017), 2307–2324. https://doi.org/10.1007/s00500-015-1942-8
71. I. Butun, B. Kantarci, M. Erol-Kantarci, Anomaly detection and privacy preservation in cloud-centric Internet of Things, in *2015 IEEE International Conference on Communication Workshop (ICCW)*, **2015** (2015), 2610–2615. https://doi.org/10.1109/ICCW.2015.7247572
72. Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, D. Breitenbacher, et al., N-baiot—network-based detection of IoT botnet attacks using deep autoencoders, *IEEE Pervas. Comput.*, **17** (2018), 12–22. https://doi.org/10.1109/MPRV.2018.03367731
73. M. A. Ferrag, L. Maglaras, DeepCoin: a novel deep learning and blockchain-based energy exchange framework for smart grids, *IEEE Trans. Eng. Manage.*, **67** (2020), 1285–1297. https://doi.org/10.1109/TEM.2019.2922936
74. H. Karimipour, A. Dehghantanha, R. M. Parizi, K. K. R. Choo, H. Leung, A deep and scalable unsupervised machine learning system for cyber-attack detection in large-scale smart grids, *IEEE Access*, **7** (2019), 80778–80788. https://doi.org/10.1109/ACCESS.2019.2920326
75. Q. Shafi, A. Basit, S. Qaisar, A. Koay, I. Welch, Fog-assisted SDN controlled framework for enduring anomaly detection in an IoT network, *IEEE Access*, **6** (2018), 73713–73723. https://doi.org/10.1109/ACCESS.2018.2884293
76. S. Prabavathy, K. Sundarakantham, S. M. Shalinie, Design of cognitive fog computing for intrusion detection in Internet of Things, *J. Commun. Netw.*, **20** (2018), 291–298. https://doi.org/10.1109/JCN.2018.000041
77. E. Anthi, L. Williams, M. Slowinska, G. Theodorakopoulos, P. Burnap, A supervised intrusion detection system for smart home IoT devices, *IEEE Internet Things J.*, **6** (2019), 9042–9053. https://doi.org/10.1109/JIOT.2019.2926365
78. V. Garcia-Font, C. Garrigues, H. Rifà-Pous, Attack classification schema for smart city WSNs, *Sensors*, **17** (2017), 771. https://doi.org/10.3390/S17040771
79. M. M. Rashid, J. Kamruzzaman, M. M. Hassan, T. Imam, S. Gordon, Cyberattacks detection in IoT-based smart city applications using machine learning techniques, *Int. J. Environ. Res. Public Health*, **17** (2020), 9347. https://doi.org/10.3390/ijerph17249347
80. R. Kozik, M. Choraś, M. Ficco, F. Palmieri, A scalable distributed machine learning approach for attack detection in edge computing environments, *J. Parallel Distrib. Comput.*, **119** (2018), 18–26. https://doi.org/10.1016/J.JPDC.2018.03.006

81. A. K. Shrivas, A. K. Dewangan, An ensemble model for classification of attacks with feature selection based on KDD99 and NSL-KDD data set, *Int. J. Comput. Appl.*, **99** (2014), 8–13. https://doi.org/10.5120/17447-5392
82. N. F. Haq, A. R. Onik, F. M. Shah, An ensemble framework of anomaly detection using hybridized feature selection approach (HFSA), in *2015 SAI Intelligent Systems Conference (IntelliSys)*, **2015** (2015), 989–995. https://doi.org/10.1109/INTELLISYS.2015.7361264
83. D. P. Gaikwad, Intrusion detection system using ensemble of rule learners and first search algorithm as feature selectors., *Int. J. Comput. Netw. Inf. Secur.*, **13** (2021), 26−34. https://doi.org/10.5815/ijcnis.2021.04.03
84. M. A. Jabbar, R. Aluvalu, S. S. S. Reddy, Cluster based ensemble classification for intrusion detection system, in *Proceedings of the 9th International Conference on Machine Learning and Computing*, **2017** (2017), 253–257. https://doi.org/10.1145/3055635.3056595
85. A. Khraisat, I. Gondal, P. Vamplew, J. Kamruzzaman, A. Alazab, A novel ensemble of hybrid intrusion detection system for detecting Internet of Things attacks, *Electronics*, **8** (2019), 1210. https://doi.org/10.3390/electronics8111210
86. A. A. Diro, N. Chilamkurti, Distributed attack detection scheme using deep learning approach for Internet of Things, *Future Gener. Comput. Syst.*, **82** (2018), 761–768. https://doi.org/10.1016/j.future.2017.08.043
87. A. Khraisat, I. Gondal, P. Vamplew, J. Kamruzzaman, A. Alazab, Hybrid intrusion detection system based on the stacking ensemble of C5 decision tree classifier and one class support vector machine, *Electronics*, **9** (2020). https://doi.org/10.3390/electronics9010173
88. M. H. L. Louk, B. A. Tama, Dual-IDS: a bagging-based gradient boosting decision tree model for network anomaly intrusion detection system, *Expert Syst. Appl.*, **213** (2023), 119030. https://doi.org/10.1016/j.eswa.2022.119030
89. S. Krishnaveni, S. Sivamohan, S. Sridhar, S. Prabhakaran, Network intrusion detection based on ensemble classification and feature selection method for cloud computing, *Concurrency Comput. Pract. Exper.*, **34** (2022), e6838. https://doi.org/10.1002/cpe.6838
90. H. Zhang, J. L. Li, X. M. Liu, C. Dong, Multi-dimensional feature fusion and stacking ensemble mechanism for network intrusion detection, *Future Gener. Comput. Syst.*, **122** (2021), 130–143. https://doi.org/10.1016/J.FUTURE.2021.03.024
91. B. A. Tama, L. Nkenyereye, S. M. R. Islam, K. S. Kwak, An enhanced anomaly detection in web traffic using a stack of classifier ensemble, *IEEE Access*, **8** (2020), 24120–24134. https://doi.org/10.1109/ACCESS.2020.2969428
92. O. A. Alghanam, W. Almobaideen, M. Saadeh, O. Adwan, An improved PIO feature selection algorithm for IoT network intrusion detection system based on ensemble learning, *Expert Syst. Appl.*, **213** (2023), 118745. https://doi.org/10.1016/j.eswa.2022.118745
93. Z. Wang, J. Liu, L. Sun, EFS-DNN: an ensemble feature selection-based deep learning approach to network intrusion detection system, *Secur. Commun. Netw.*, **2022** (2022), 2693948. https://doi.org/10.1155/2022/2693948
94. A. Nazir, R. A. Khan, A novel combinatorial optimization based feature selection method for network intrusion detection, *Comput. Secur.*, **102** (2021), 102164. https://doi.org/10.1016/j.cose.2020.102164.

95. B. A. Tama, M. Comuzzi, K. H. Rhee, TSE-IDS: a two-stage classifier ensemble for intelligent anomaly-based intrusion detection system, *IEEE Access*, **7** (2019), 94497–94507. https://doi.org/10.1109/ACCESS.2019.2928048.
96. S. Rajagopal, P. P. Kundapur, K. S. Hareesha, A stacking ensemble for network intrusion detection using heterogeneous datasets, *Secur. Commun. Netw.*, **2020** (2020). https://doi.org/10.1155/2020/4586875
97. F. Salo, A. B. Nassif, A. Essex, Dimensionality reduction with IG-PCA and ensemble classifier for network intrusion detection, *Comput. Netw.*, **148** (2019), 164–175. https://doi.org/10.1016/j.comnet.2018.11.010
98. B. A. Tama, K. H. Rhee, An extensive empirical evaluation of classifier ensembles for intrusion detection task, *Comput. Syst. Sci. Eng.*, **2** (2017), 149–158.
99. N. Acharya, S. Singh, An IWD-based feature selection method for intrusion detection system, *Soft Comput.*, **22** (2018), 4407–4416. https://doi.org/10.1007/s00500-017-2635-2
100. B. Selvakumar, K. Muneeswaran, Firefly algorithm based feature selection for network intrusion detection, *Comput. Secur.*, **81** (2019), 148–155. https://doi.org/10.1016/J.COSE.2018.11.005
101. H. Alazzam, A. Sharieh, K. E. Sabri, A feature selection algorithm for intrusion detection system based on Pigeon Inspired Optimizer, *Expert Syst. Appl.*, **148** (2020), 113249. https://doi.org/10.1016/J.ESWA.2020.113249
102. A. S. Eesa, Z. Orman, A. M. A. Brifcani, A new feature selection model based on ID3 and bees algorithm for intrusion detection system, *Turk. J. Electr. Eng. Comput. Sci.*, **23** (2015), 615–622. https://doi.org/10.3906/ELK-1302-53
103. T. A. J. Ali, M. Jawhar, Proposing a model for detecting intrusion network attacks using machine learning techniques, *J. Educ. Sci.*, **31** (2022), 99–109. https://doi.org/10.33899/edusj.2022.133867.1240
104. Y. Deng, H. Duan, Control parameter design for automatic carrier landing system via pigeon-inspired optimization, *Nonlinear Dyn.*, **85** (2016), 97–106. https://doi.org/10.1007/S11071-016-2670-Z
105. T. Guilford, S. Roberts, D. Biro, I. Rezek, Positional entropy during pigeon homing II: navigational interpretation of Bayesian latent state models, *J. Theor. Biol.*, **227** (2004), 25–38. https://doi.org/10.1016/j.jtbi.2003.07.003
106. J. Kennedy, R. C. Eberhart, A discrete binary version of the particle swarm algorithm, in *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, IEEE, (1997), 4104–4108. https://doi.org/10.1109/ICSMC.1997.637339
107. V. Sugumaran, V. Muralidharan, K. I. Ramachandran, Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing, *Mech. Syst. Signal Process.*, **21** (2007), 930–942. https://doi.org/10.1016/J.YMSSP.2006.05.004
108. M. Abdulrazaq, A. Salih, Combination of multi classification algorithms for intrusion detection system, *Int. J. Sci. Eng. Res.*, **6** (2015), 1364–1371.
109. Q. Zhang, Y. Qu, A. Deng, Network intrusion detection using kernel-based fuzzy-rough feature selection, in *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, **2018** (2018). https://doi.org/10.1109/FUZZ-IEEE.2018.8491578
110. P. S. Varma, V. Anand, Random Forest learning based indoor localization as an IoT service for smart buildings, *Wireless Pers. Commun.*, **117** (2021), 3209–3227. https://doi.org/10.1007/s11277-020-07977-w

111. Y. Amit, D. Geman, Shape quantization and recognition with randomized trees, *Neural Comput.*, **9** (1997), 1545–1588. https://doi.org/10.1162/neco.1997.9.7.1545
112. S. S. Dhaliwal, A. A. Nahid, R. Abbas, Effective intrusion detection system using XGBoost, *Information*, **9** (2018), 149. https://doi.org/10.3390/info9070149
113. I. Sharafaldin, A. H. Lashkari, S. Hakak, A. A. Ghorbani, Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy, in *2019 International Carnahan Conference on Security Technology (ICCST)*, (2019), 1–8. https://doi.org/10.1109/CCST.2019.8888419
114. A. K. Koliopoulos, P. Yiapanis, F. Tekiner, G. Nenadic, J. Keane, A parallel distributed Weka framework for big data mining using spark, in *2015 IEEE International Congress on Big Data*, (2015), 9–16. https://doi.org/10.1109/BigDataCongress.2015.12
115. M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauly, et al., Resilient distributed datasets: a fault-tolerant abstraction for in-memory cluster computing, in *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*, (2012), 15–28.
116. M. Hall, Advanced data mining with Weka, Online Course, University of Waikato, 2016
117. W. Li, Z. Liu, A method of SVM with normalization in intrusion detection, *Procedia Environ. Sci.*, **11** (2011), 256–262. https://doi.org/10.1016/j.proenv.2011.12.040
118. Scikit-learn developers, 2022. Available from: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.htm.
119. D. H. Wolpert, Stacked generalization, *Neural Netw.*, **5** (1992), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1

Appendix

This section provides instructions for setting up Weka-Spark to work with a Hadoop cluster in order to continue managing data for the intrusion detection issue in cloud networks for smart cities. Weka's ability to be used as a data mining tool is one of its benefits. Since it was created in Java, it can function on any OS as long as JVM is installed (Java Virtual Machine). Applications written for Spark can operate on any OS, albeit Windows requires some specialized libraries for execution whereas Linux does not. To use the Spark utilities in Weka, in addition to these prerequisites, a number of basic setups must be completed. First, the environment variables for Java and Hadoop must be established. Second, the Hadoop environment variable's path must contain both the hadoop.dll and the winutils library, and at last DistributedWekaBase and DistributedWekaSpark dependencies need to be installed on the Weka. With all of the prior Weka Knowledge Flow settings, a result similar to that in Figure A.1 is shown.

The tool WekaClassifierEvaluationSparkJob is used for J48, RF, and XGBoost training and evaluation. It allows assessing a classifier using cross-validation, a distinct dataset, or training data. In our case, for training, the training dataset is used while for testing the data, a distinct test dataset is employed. The data for training is passed through the ArffHeaderSparkJob, so some of the most crucial variables must be indicated, including the path (Copy the datasets' CSV files to HDFS and point the inputFile setting of the ArffHeaderSparkJob to the HDFS location) to the data to execute the evaluation and the chosen classifier, stacking. Three base classifiers are added to the stacking section's classifier field, and XGBoost is selected as the meta-classifier. The fields that contain the route in HDFS to a different test suite that includes the path detail to the test dataset are some of the most crucial fields.

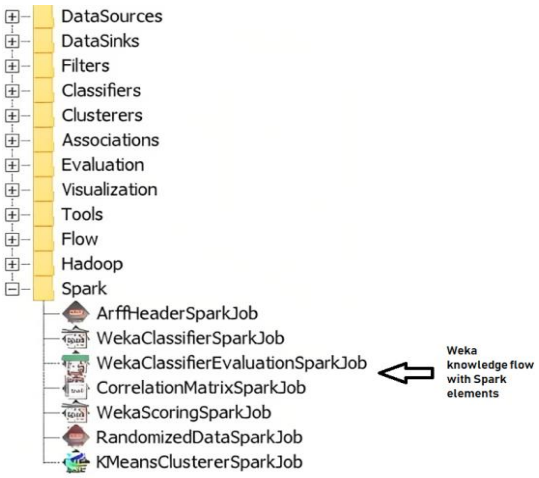


Figure A.1. Spark elements in the Weka's knowledge flow palette.

Weka's KnowledgeFlow is an alternative to Weka Explorer with a graphical environment. It is well known for its simplicity of use because it is a fairly intuitive work system with a graphical interface that allows you to drag the objects from a palette to the workspace and create connections between them in various types to obtain the results and information.

All of the filters, classifiers, regressors, and other tools that are present in the version of Weka that is being used can be used in KnowledgeFlow [117]. Additionally, some additional tools can be used, such as in the case presented in this work, which is the distributed processing capability of Hadoop and Spark. While Explorer only handles batch data, KnowledgeFlow can process data incrementally or in large batches [117]. Once the libraries are installed and configured, one can find the many Spark jobs that can be carried out in the Weka component palette. The TextViewer is the last visualization tool that is suggested; it enables the data to be obtained in text form for subsequent analysis and storage. Once the runs are completed, getting the results is quite easy because they are stored in the output configuration folder, 'OutputDir', as well as in the TextViewer, which allows you to view the results of the tests.

In addition to Weka's KnowledgeFlow, a Web service offered by Spark that is active on the computer where the application driver is running allows for more detailed monitoring of the applications and their progress. With access through a Web browser to the website where the cluster jobs are monitored through its URL and its web access port, by default this is 8080, you can check if the program is discovered running, as well as the resources it has, in addition to other extensive capabilities.

AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)

Sādhanā (2024)49:62
https://doi.org/10.1007/s12046-023-02418-3

© Indian Academy of Sciences

Soil-slope stability investigation using different nail inclinations: a comprehensive LSD, FEM and experimental approach

PRASHANT C RAMTEKE* and ANIL KUMAR SAHU

Civil Engineering Department, Delhi Technological University, New Delhi 110 042, India
e-mail: prashant_2k16phdce12@dtu.ac.in; anilsahu@dtu.ac.in

MS received 16 June 2023; revised 19 November 2023; accepted 22 November 2023

Abstract. Soil nailing is a versatile construction technique used for retaining structures, offshore structures, structures rehabilitation, and stabilizing natural as well as man-made earth slopes. This method was initially evolved in Europe and recognized in various provinces. This research investigates the relationship between uniform vertical and uniform horizontal loading on soil slopes to gain new insights into the effectiveness of soil nailing in stabilizing earth slopes. This study used a novel numerical investigation to determine the maximum factor of safety (FOS) at various nail inclinations of 0°, 5°, 10°, 15°, 20°, and 25° with the horizontal plane. To evaluate the effectiveness of the soil nail system for slope stabilization, the study considered significant limiting factors by using the limit state design (LSD) approach. In addition, the numerical method for identifying slope displacement, mathematical modelling for identifying nail displacement within soil slope, and experimental investigation in the laboratory were carried out in this study. The investigation's findings show that soil nailing is the more effective technique for slope stabilization. The 15° nail inclination with the horizontal plane produced the best results in terms of stability and FOS. Furthermore, considering the critical limiting factors, an inclination of 15° for soil nails exhibits optimal effectiveness in strengthening the stability of a vertical cut soil slope. This inclination angle allows for efficient load transfer and distribution and decreases the risk of slope failure. Overall, the 15° nail inclination with the horizontal plane provides valuable insights and an increased FOS.

Keywords. Factor of safety; limit states design; reinforced soil; soil nailing; slope settlement; slope stabilization.

1. Introduction

Soil nailing is a construction technique that makes existing slopes stronger and more stable [1]. It is commonly used for slope stabilization in highways and railways. The construction of the soil nailing process starts from the top to the bottom. In most cases, this technique is more cost-effective than some other slope stabilization techniques [2–4]. Due to the economic benefits and advantages associated with soil nailing, nowadays this technique is more popular and has produced rapid development of nailing technology. Soil nails reproduce their reinforcing action by soil deformation [5–8]. In the soil nails, tensile forces are developed due to deformations, and most of the resistance gives rise to tension [9–12]. There are various ground improvement and slope stabilization techniques available to increase soil quality and these methods can be categorized in several ways. The soil reinforcement, compaction, vertical drainage, electro-osmosis, thermal stabilization, ion exchange, and soil stabilization by grouting, methods were suggested by Mitchell [13]. Similarly, Munfakh and Wyllie [14] proposed eight main groups for slope stabilization such as reinforcement, consolidation, weight reduction, densification, thermal stabilization, biotechnical stabilization, electro-treatment, and chemical treatment etc. The experimental studies described by Jewell and Pedley [15] in their most prominent and broad research explored soil nail bending stiffness and the effect of displacement.

Various methods are available for designing the soil-nailed structures on the limit equilibrium method (LEM) concept. Due to the reduced number of required parameters and their simplicity, most of the methods are popular. However, they are not capable of anticipating soil movements. During soil nailing, to develop the maximum tensile force, the reinforcements are installed horizontally as well as slightly inclined parallel to the direction of tensile stresses. Soil nailed slope improves soil shearing resistance because of passive inclusions [16]. The soil nailing structures get split into the active zone and passive zone (refer to figure 1). The active zone is deformed first, and at the time of slope failure, the tensile forces are developed in the soil nails in the active and passive zones. This tensile force

*For correspondence
Published online: 10 February 2024

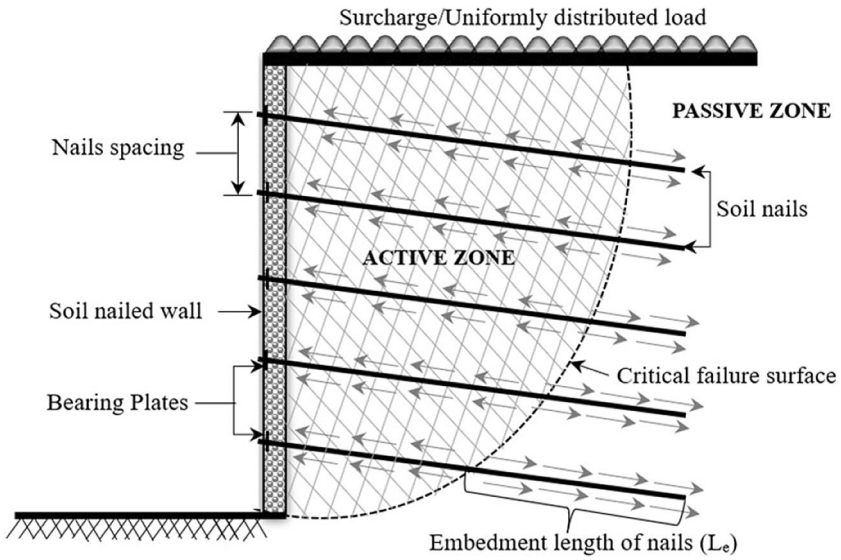


Figure 1. Emblematic cross-section of a soil-nailed slope.

decreases the applied shear force (SF) while increasing the normal force on the slip plane [17]. Soil nails are embedded in the passive zone, hence resistance between the nails and soil prevents the departure of the nails from the slope. Based on this principle, adequate nail length should be provided in the counter-active zone. Moreover, the combined efficiency of tensile force is generated in the active zone. For the identification of soil nailed slope behaviour, a two-dimensional (2-D) simulation model was used by Cheuk *et al* [18] and Fan and Luo [7] in their study.

Particularly Fan and Luo's [7] study concentrated on optimizing soil nail arrangements for slope stabilization, considering nail orientation and layout using a non-linear finite element method to evaluate stability through the factor of safety (FOS). They discovered that the ideal nail orientation varies depending on slope angle and backslope gradient, underscoring the need for personalized strategies for specific slope conditions. Similarly, Jelušič and Žlender [19] introduced a safety-factor optimization method for soil-nail walls using non-linear programming (NLP). They developed ANFIS models (ANFIS-INC and ANFIS-SF) to estimate optimal nail inclination and safety factors for various wall designs, emphasizing the importance of adjusting nail inclination based on design parameters. Similarly, Arvin *et al* [20] combined the FHWA manual's limit equilibrium (LEM)-based approach and nonlinear programming to optimize soil nail wall design. They determined the ideal nail inclination angle for maximum safety against failure while considering factors like nail diameter, length, soil friction, slope angle, back slope angle, and nail layout etc. Their findings highlighted that increasing nail diameter and nail length significantly enhanced slope stability, with a preference for placing nails in the lower 1/3 of the slope for optimal effectiveness in soil nail walls.

Generally, the soil nailing technique is more useful in roadway or highway cut excavations and nowadays, it is commonly used in road widening beneath existing bridges and existing retaining structures, as well as in the repair and reconstruction of retaining structures, hybrid soil nailed walls or mechanically stabilized earth (MSE) walls, and shored mechanically stabilized earth (SMSE) etc. [5, 21] This technique predominantly resists tensile forces alongside the shear force (SF) and bending moment (BM). Many researchers, such as Sivakumar Babu and Jaladurgam [22]; and Sivakumar Babu and Singh [23], have studied the effects of SF and BM on the soil nailed wall and their large deformation in the structure.

From the above literature, it is found that more researchers focused on finding the various forces and FOS under surcharge loading by the limit equilibrium method (LEM), finite element method (FEM), finite discrete element method (FDM), federal highway administration (FHWA) methods etc. Based on the available literature, it is observed that the investigation using limit states design (LSD) considering Indian Standard Code IS:456-2000 [24] specification for soil nailed wall design has not been conducted to ascertain a particular angle of nail inclination at which the vertical slope gives better efficiency or FOS. To bridge this research gap, the numerical analyses were conducted using GEO-5 (2018) fine software [25] and validated the results through FEM and laboratory experimentation. Nail inclinations ranging from 0° to 25° with the horizontal plane were considered to determine the optimal nail inclinations that maximize strength, stability, and FOS by LSD. The FHWA manual [5] stated two separate limiting factors i.e., Strength Limit States and Service Limit

States. As per the FHWA manual recommendations, these factors must be taken into account during the study and design of soil nail slopes. In this research, by considering these important limiting factors the soil nail slope has been deliberated by the LSD for the model illustrated in figure 2, which is more applicable in roadways, railways and highway construction developments. The properties of soil and nails used in this analysis are revealed in table 1.

1.1 *Scope and objective*

The scope of this research delves into the impact of nail inclination angles on soil slope stability and deformation characteristics etc. It comprehensively assesses the performance of soil nailing techniques by employing a novel numerical investigation technique, considering nail inclinations of 0°, 5°, 10°, 15°, 20°, and 25° with the horizontal plane. The study employs the Limit State Design (LSD) approach to evaluate the soil nail system's effectiveness in slope stabilization by factoring in crucial limiting parameters. Nowadays, this essential study addresses the current demand for slope stabilization in various civil engineering infrastructure projects like highways, railways etc.

The main objective of this investigation is to assess the efficacy of soil nailing in improving slope stability, with a specific emphasis on identifying the optimal nail inclination angle. To be more precise, the basic objectives incorporated are: (i) Investigating the correlation between nail inclination and slope stability, along with deformation under surcharge loading conditions. (ii) Assessing the impact of nail inclination on the FOS for soil-slope systems. Moreover, the study involves a comparative analysis of findings obtained through the LSD method, numerical simulations (FEM) and experimental investigations.

2. Methodology

This study investigates the impact of soil nails on soil slopes and their role in enhancing structural stability. The study assessed soil nail effects, analyzed force distributions, identified critical stress points, and used mathematical modelling to find nail displacement in the soil mass. The study also characterized settlement patterns within soil slopes under surcharge loading. Furthermore, the study emphasized the crucial role of experimental investigations. These methodological approaches provide a comprehensive understanding of soil nail slope behaviour and its implications in the geotechnical engineering structure as illustrated in figure 2.

2.1 *Basic principles of soil slope behaviour under uniform vertical and horizontal loading*

Uniform vertical loading primarily influences the vertical stress and settlement characteristics of a soil mass, while uniform horizontal loading introduces lateral pressures, movements, and the potential for soil arching effects [26]. The specific effects depend on factors such as the magnitude of the surcharge load, soil properties, and the

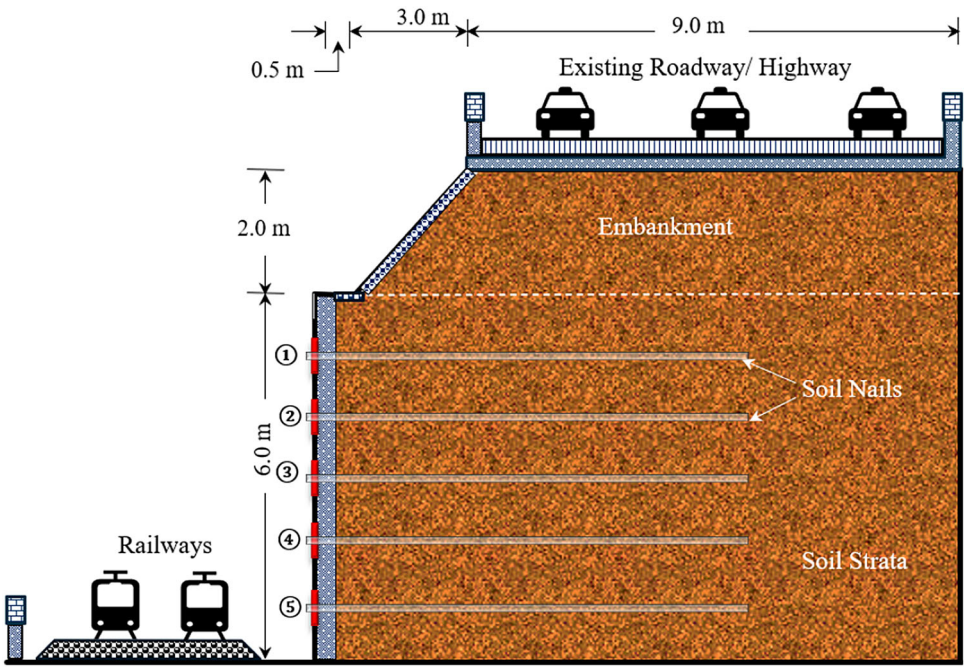


Figure 2. Profile diagram of soil nailed wall.

Table 1. Properties of soil and nails within the soil slope.

Soil properties	
Type of soil	Poorly graded sand (SP)
Model	Mohr–Coulomb
Surcharge load (w)	18.00 kN/m^2
Unit weight (γ)	19.50 kN/m^3
The angle of internal friction (φ_{ef})	27°
Saturated unit weight (γ_{sat})	19.50 kN/m^3
The cohesion of soil (c_{ef})	12.00 kPa
Poisson's ratio (ν)	0.35
Dilation angle (ψ)	0.00
Nails properties	
Nail Type	HYSD (high yielding strength deformed)
Length (L)	9.00 m
Nails spacing (S_v and S_h)	1.00 m
Diameter (ϕ)s	20 mm
Modulus of elasticity (E)	2×10^5 N/mm^2
Minimum yield stress (f_y)	500 N/mm^2
Poisson's ratio (ν)	0.30

surrounding conditions etc. A thorough comprehension of these effects is crucial for ensuring the safety and stability of construction and engineering activities in slope stability analysis.

2.1a *Stresses development within soil slope due to uniform vertical loading:* On the surface of the semi-infinite elastic mass vertical load of the Q/unit area acts uniformly on a flexible infinite strip shown in figure 3. The width δx is an elementary strip considered at a distance x from the centreline of the uniform load. Q.dx is the load per unit length of the elementary strip, it can be considered like line load. Due to this elementary strip loading, the increase of vertical stresses (σ_{zv}) at point (P) is obtained by using the following equations [27] :

To evaluate the increase in vertical stress ($\sigma_{z,v}$) at a specified point (P) within a soil slope induced by a uniformly applied vertical load (Q_v), Equation (1) is employed. This equation integrates both the orientation angle (α) and the angle of application (δ), providing valuable insights into the distribution of vertical stress. By utilizing this equation, we have ascertained the magnitude of vertical stress ($\sigma_{z,v}$) within the soil slope resulting from uniform vertical loading (Q_v).

$$\sigma_{z,v} = \frac{Q_V}{\pi}[\alpha + \sin\alpha\cos(\alpha + 2\delta)] \tag{1}$$

where $\sigma_{z,v}$ is vertical stress due to uniform vertical loading, Q_v is the vertical loading, α is the orientation angle and δ is the angle of application.

Equation (2) defines the impact of vertical loading (Q_v) on the development of horizontal stress ($\sigma_{x,v}$) in the soil slope. It supports equation (1) by taking into account the horizontal stress response in the slope. This equation helps to determine the horizontal stresses that arise due to uniform vertical loading.

$$\sigma_{x,v} = \frac{Q_V}{\pi}[\alpha - \sin\alpha\cos(\alpha + 2\delta)] \tag{2}$$

where $\sigma_{x,v}$ is horizontal stress due to uniform vertical loading, Q_v is the vertical loading, α is the orientation angle and δ is the angle of application.

Equation (3) is used to evaluate the shear stress ($\tau_{(XZ),v}$) on the vertical plane of the soil slope. This is a critical parameter for understanding potential sliding events and shear failure. The shear stress due to uniform vertical loading within the soil slope has been calculated by the given equation:

$$\tau_{(XZ),v} = \frac{Q_V}{\pi}[\sin\alpha\sin(\alpha + 2\delta)] \tag{3}$$

where $\tau_{(XZ),v}$ is shear stress due to uniform vertical loading, Q_v is the vertical loading, α is the orientation angle and δ is the angle of application.

2.1b *Stresses development within soil slope due to uniform horizontal loading:* The stresses at a point (P) on an infinite strip, after a uniform horizontal load is applied, can be determined using the same procedure used for vertical loading. Q.dx is the load per unit length, dx is the width of the elementary strip shown in figure 3.

Equation (4) is used to estimate the vertical stress at point (P) due to the application of uniform horizontal loading (Q_H). It indicates the change in vertical stress ($\sigma_{z,\ H}$) at the point (P). Uniform horizontal loading (Q_H) represents the magnitude of the horizontal load applied to the surface of the soil. This loading introduces lateral forces into the soil mass. Corresponding to the previous equations, α is the orientation angle. It plays a crucial role in determining the response of the soil to horizontal loading. δ represents the angle of application at which the horizontal load is applied to the soil. To obtain the vertical stress ($\sigma_{z,H}$) under horizontal loading (Q_H), we used the following equation (4)

$$\sigma_{z,H} = \frac{Q_H}{\pi}[\sin\alpha\sin(\alpha + 2\delta)] \tag{4}$$

where $\sigma_{z,H}$ is vertical stress due to uniform horizontal loading, Q_H is the horizontal loading, α is the orientation angle and δ is the angle of application.

To evaluate the horizontal stress ($\sigma_{x,H}$) generated by lateral forces within the soil mass, equation (5) is used. The values of R_1 and R_2, which are defined in Equation 5 and figure 3, are determined based on the distance between point (P) and the boundaries of the loaded length/area. These values strongly impact the distribution of horizontal stresses. In this research, the AutoCAD software tool is used for obtaining these R_1 and R_2 measurements. When the soil mass is subjected to uniform horizontal loading (Q_H), the horizontal stress ($\sigma_{x,H}$) is

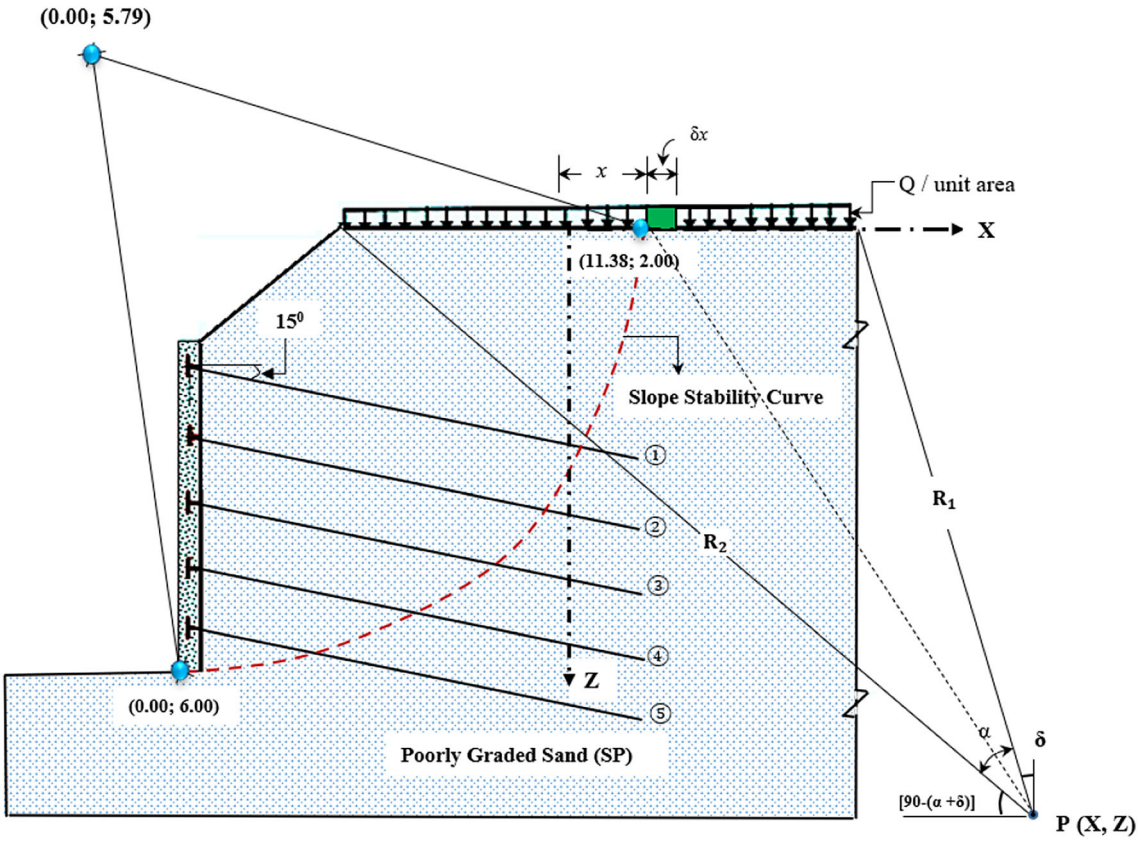


Figure 3. The geometrical figure for optimization of soil nailed slope design.

$$\sigma_{x,H} = \frac{Q_H}{\pi}\left[\left[2.303\log\left(\frac{R_1}{R_2}\right)^2\right] - \sin\alpha\sin(\alpha+2\delta)\right] \quad (5)$$

where $\sigma_{x,H}$ is horizontal stresses due to uniform horizontal loading, Q_H is the horizontal loading, α is the orientation angle and δ is the angle of application.

To compute the shear stresses ($\tau_{(XZ),H}$) on the horizontal plane within the soil caused by horizontal loading, we used equation (6). These stresses indicate the likelihood of sliding or deformation within the soil mass/slope. The equation (6) helps us determine the shear stress that arises from uniform horizontal loading.

$$\tau_{(XZ),H} = \frac{Q_H}{\pi}[\alpha - \sin\alpha\cos(\alpha+2\delta)] \quad (6)$$

where $\tau_{(XZ),H}$ is shear stress due to uniform horizontal loading, Q_H is the horizontal loading, α is the orientation angle and δ is the angle of application.

2.2 *Comparative analysis between effects of uniform vertical loading and horizontal loading*

Uniform vertical loading (Q_v) and uniform horizontal loading (Q_H) are two types of external forces that can be applied to a structure or material. Table 2 reveals the key differences between these two loading conditions. Let's analyze the results to gain a better understanding of their distinctions.

2.2a *Effects uniform vertical loading (Q_v) within the soil slope:* When a vertical load (i.e. downward pressure), is applied to the structure, the vertical stress ($\sigma_{z,v}$) increases proportionally to the extended loaded length. For example, at a loaded length of 5 m, the vertical stress is 74.417 kN/m^2, and it goes up to 166.783 kN/m^2 at a loaded length of 9 m. This means the structure experiences more vertical compression with increasing length. Additionally, the indirect stress ($\sigma_{x,v}$) also increases from 74.552 to 165.533 kN/m^2, indicating a tendency for the structure to deform horizontally. The shear stress in the plane ($\tau_{(XZ),v}$) increases from 1.287 to 2.707 kN/m^2, suggesting some twisting or shearing forces within the soil slope.

2.2b *Effects of uniform horizontal loading (Q_H) within the soil slope:* When a horizontal force is applied to the structure, the vertical stress ($\sigma_{z,H}$) remains relatively constant with the loaded length, indicating minimal vertical compression. At a loaded length of 5 m, it is 1.287 kN/m^2, and at 9 m, it remains 2.707 kN/m^2. However, the horizontal stress ($\sigma_{x,H}$) increases from − 4.121 kN/m^2 (tension) to − 7.592 kN/m^2 (tension) with increasing length.

This means the structure experiences tension in the horizontal direction as it gets longer. The shear stress in the plane ($\tau_{(XZ),\ H}$) due to horizontal loading increases from 74.552 to 165.533 kN/m^2, indicating a tendency for the structure to twist or shear horizontally.

In summary, vertical loading (Q_v) leads to increased compression vertically, along with some twisting or shearing forces. Horizontal loading (Q_H) primarily induces tension in the horizontal direction, with minimal vertical compression, and also causes some twisting or shearing forces horizontally. These numerical values help us understand how different types of loading affect the stresses and deformations within the structure. The results from table 2 and the previous discussion show that uniform horizontal loading causes higher shear stresses ($\tau_{(XZ),H}$) than uniform vertical loading. Shear stresses can significantly affect the stability of soil slopes and can lead to structural failures or landslides. Using soil nails within the soil slope we can reduce these risks and improve slope stability performance [3, 28–32]. The impact of soil nailing with different nail inclinations on enhancing soil slope stability is investigated and discussed in underneath.

3. Result and discussion

3.1 *The influence of nail inclinations within the soil slope*

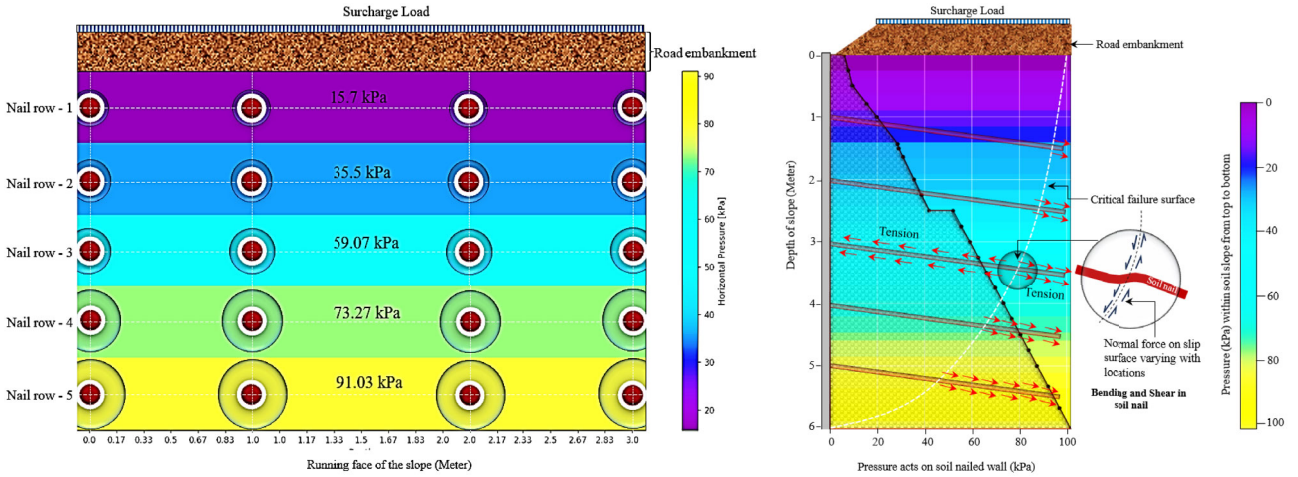
To ensure the stability of the soil slope and the effectiveness of soil-nailing structures detailed analysis has been carried out considering various factors. This includes examining pressure and nail force variations within soil slope, SF and BM considerations, FOS and slip surface resistance assessments, mathematical modelling for displacement predictions, and gaining a comprehensive understanding of the internal force distribution along the soil nailing structure. All of these basics are vital in optimizing the stability of the soil slope and enhancing the effectiveness of soil-nailing structures.

3.1a *Observation of nail forces under different nail inclinations:* Under the surcharge load, the horizontal and vertical pressures are developed within the soil slope and it move towards the soil-nailed wall (refer to figure 4). During this phenomenon, axial tensile forces are produced in all the nails (figure 4b). Whereas due to vertical loading large bending moments take place in the soil nails and the structure reaches collapse due to large displacements within the soil slope and structure cannot be serviceable [33]. Considering this concept in mind, the analysis was performed by using Geo5 fine software to observe the effect of different nail inclinations within the soil slope. Results of the analysis show that the pressure around the nails and nail forces increase as we move towards the lowermost nails from nail 1 to nail 5 (refer to figurse 4a and 5). Figure 5 shows the observed nail forces at different nail inclinations. However, there is not much variation in nail forces in spite of nail inclination up to 25°. The consistent and minimal variation in nail forces observed across different inclinations i.e. for 0°, 5°, 10°, 15°, 20°, and 25° with the horizontal plane, suggests potential symmetrical or well-distributed load/pressure sharing characteristics within the structural system (refer to figure 5). This finding implies that the influence of nail inclination on force transmission is relatively minor in the context of nail forces at various inclinations. The uniformity in forces is influenced by the diameter and length of the soil nails as well as the structure's geometry, load distribution, material properties etc.

3.1b *Forces on slip surface resist for soil and nails:* The term "Forces on slip surface resistance for soil and nails" refers to the analysis of resistant forces acting on/or near the critical slip surface within soil mass, particularly in scenarios involving slope stability or retaining structures. These forces include both internal factors, such as the shear strength and cohesion within the soil, and external influences like applied loads and geometry of the slope etc. In the present study, it was observed that at a nail inclination of 15°, the soil on the slip surface resists a maximum force of 402.27 kN/m, while the nails, undergoing deformation, resist a maximum force of 198.65 kN/m. The intricate

Table 2. Stresses developed due to uniform vertical loading and uniform horizontal loading.

	For uniform vertical loading (Q_v)					For uniform horizontal loading (Q_H)				
Loaded length (m)	α	δ	$\sigma_{z,v}$ (kN/m^2)	$\sigma_{x,v}$ (kN/m^2)	$\tau_{(XZ),v}$ (kN/m^2)	R_1 (m)	R_2 (m)	$\sigma_{z,H}$ (kN/m^2)	$\sigma_{x,H}$ (kN/m^2)	$\tau_{(XZ),H}$(kN/m^2)
5	13°	40°	74.417	74.552	1.287	13.120	16.800	1.287	−4.121	74.552
6	17°	37°	97.374	97.432	1.675	12.500	16.800	1.675	−5.063	97.432
7	20°	33°	114.728	114.455	1.955	11.930	16.800	1.955	−5.878	114.455
8	25°	29°	143.535	142.944	2.403	11.410	16.800	2.403	−6.838	142.944
9	29°	24°	166.783	165.533	2.707	10.970	16.800	2.707	−7.592	165.533



(a) Horizontal pressure within soil slope **(b)** Soil nailed slope behaviour under vertical loading

Figure 4. Identifying the effects of different loads on soil nailed slopes.

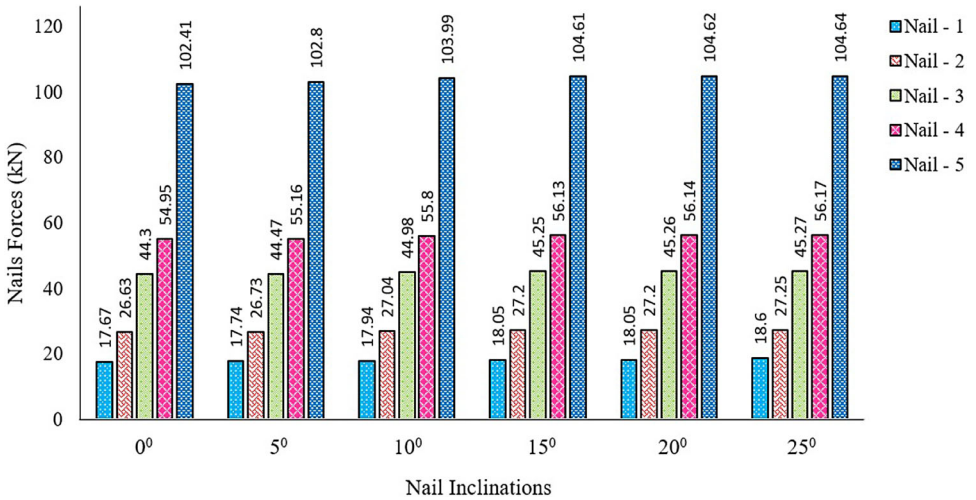


Figure 5. Nails forces at various nail inclinations.

relationship between nail inclinations and the corresponding forces acting on the slip surface, resisted jointly by the soil and the reinforcing nails, is vividly depicted in figures 6 and 7. This graphical representation provides valuable insight into the interplay between the orientation of the nails and the forces they counteract on/near the critical slip surface, providing crucial information for optimizing slope stabilization strategies.

3.1c *Maximum SF and BM in nail locations:* Under a small surcharge loading, the BM did not procreate prominent SF perpendicular to the nail as compared to the axial forces. However, under the large surcharge loading, higher SFs are generated in nails. The mobilized SF within the nails increased from top to bottom. Under the service loading, the roles of SF and BM are different. Analytically the distribution of SF along each nail has been observed and the maximum SF value i.e., 45.52 kN is found at the lowermost nail location, i.e., at nail number 5 (refer to figure 8). Also, It is important to consider various factors, such as BM, when assessing structural integrity. Despite their small magnitude in comparison to SF and axial tensile forces (P), BM is significant in ensuring overall stability. Furthermore, when subjected to a surcharge load, the entire structure experiences straining. A study of BM at various nail locations revealed that, under surcharge loading, the lowermost nail location had the highest BM of 7.59 kN m (refer to figure 8). These findings provide valuable insights into structural behaviour, which helps create stronger and safer structures.

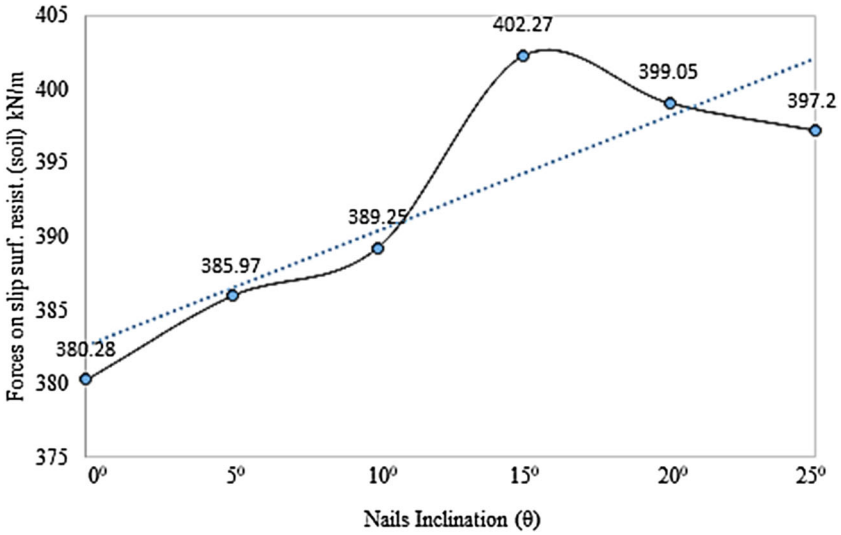


Figure 6. Force on slip surface resisted by soil mass.

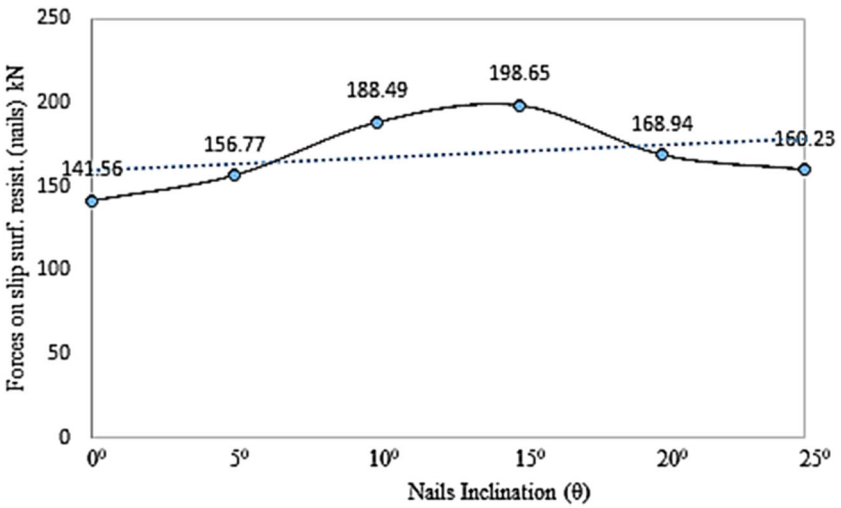


Figure 7. Force on slip surface resisted by soil nails.

3.2 *Impact of nail inclinations on safety factor*

The analysis was carried out using the soil slope model shown in figure 2 to obtain the FOS. The output of the relationship between the calculated FOS and nail inclinations for the given soil slope model is shown in figure 9. From the output (figure9) it is observed that the FOS consistently increased within the range of nail inclinations from 0° to 15°. However, beyond 15°, there is a sudden decrease in FOS, reflecting a reduction in the reinforcing action of the nails with increasing inclination. The maximum FOS of 1.630 is observed at 15° nail inclinations. At 25°, the FoS is notably reduced to 1.573, as shown in figure 9. This indicates that soil nails at steep inclinations reduced the FOS and affected soil slope stability. Additionally, in section 3.1b the analysis revealed that the highest resisting forces on the failure slip surface for both soil and nails are observed at a 15° inclination (refer to figures 6 and 7). Therefore, in the present analysis, it is observed that soil nails at a 15° inclination are most effective in resisting slope failure.

4. Mathematical modelling for predicting nail displacement in soil mass

In this study, the forces on the slip surface resisted by soil and nails are computed (figures 6 and 7). It is observed that near the critical failure surface, the forces in the soil mass

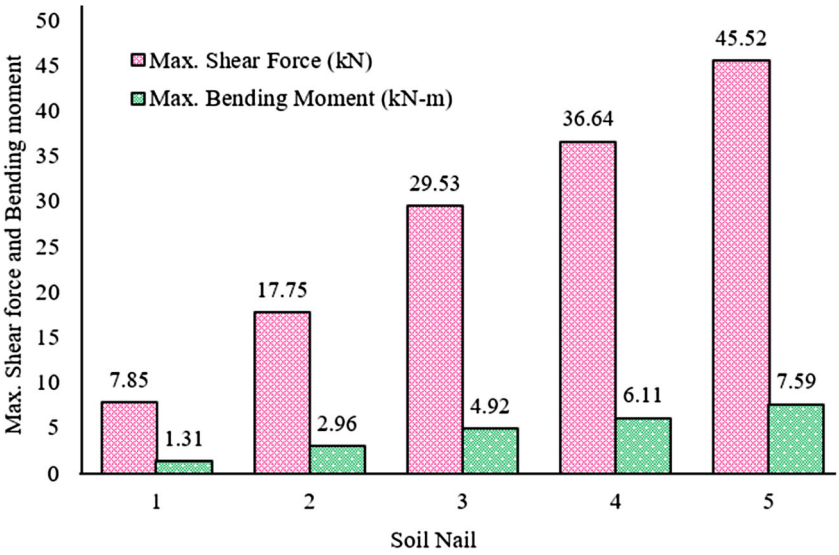


Figure 8. Maximum shear force (SF) and bending moment (BM) at different nails.

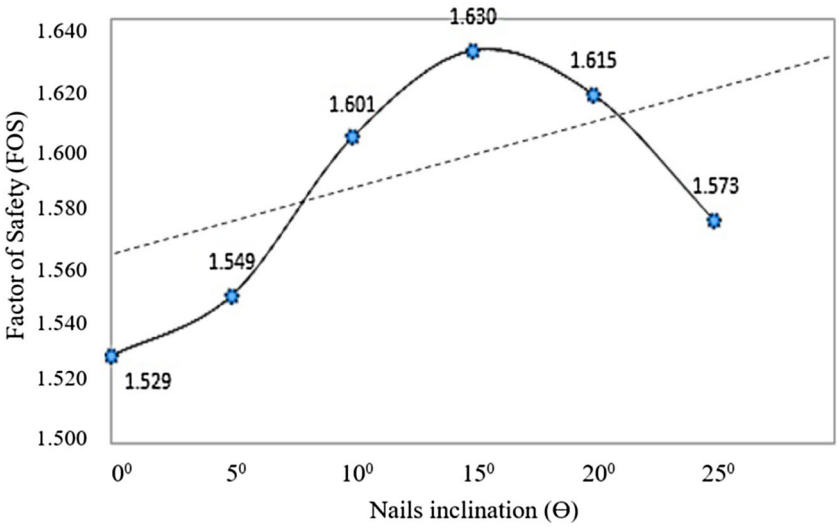


Figure 9. The factor of safety (FOS) for different angles of nails inclination.

lead to the displacement of nails in the active and passive zones (i.e. plastic zone and elastic Zone). This affects the overall structure's stability as the reinforcement does not get distributed evenly in the entire soil mass. To understand the displacement of nails in the soil mass, a predictable linear spring system model is used which consists of two spring systems connected in series with stiffness k_1 and k_2 (figure 10(a–c)). A force P acts in the x-direction (i.e. +ve) on the right-hand side, while the left-hand side is rigidly fixed. By considering this mathematical model, the displacement of the reinforcement in the soil mass at different nail inclinations is obtained.

4.1 *Global stiffness matrix*

4.1a *Establishment of element stiffness matrix for finding displacement in soil-nail mass:* Figure 10(c) shows a standard single-degree freedom system with two nodes, u_2 and u_3, each of these has a potential displacement. The spring component of this system has a stiffness of 'k' and only one direction of displacement is possible, i.e., x-direction. Using the theory of force equilibrium and taking into account the force on the slip surface resisted by soil (k_1) and the slip surface resisted by nails (k_2), a mathematical model has been developed. The model presented in this study allows for the formulation of the element

stiffness matrix of elements. Specifically, the element stiffness matrix can be expressed as follows:

$$[k_1] = \begin{bmatrix} k_1 & -k_1 \\ -k_1 & k_1 \end{bmatrix}, \text{ and } [k_2] = \begin{bmatrix} k_2 & -k_2 \\ -k_2 & k_2 \end{bmatrix} \quad (7)$$

In a global format, the element stiffness matrix for (k_1) and (k_2) is placed as.

$$[k_1] = \begin{bmatrix} k_1 & -k_1 & 0 \\ -k_1 & k_1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \text{ and } [k_2] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & k_2 & -k_2 \\ 0 & -k_2 & k_2 \end{bmatrix} \quad (8)$$

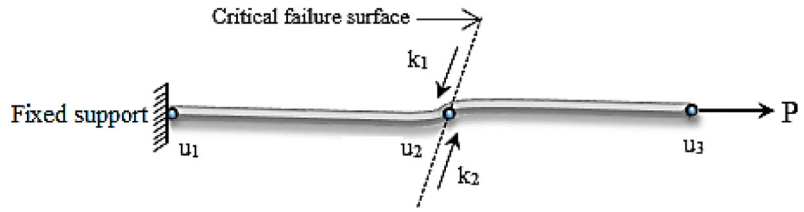
A global stiffness matrix [K] is formed by combining matrix [K_1] and matrix [K_2].

$$\begin{aligned} [K] &= [K_1 + K_2] \\ &= \begin{bmatrix} k_1 & -k_1 & 0 \\ -k_1 & k_1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & k_2 & -k_2 \\ 0 & -k_2 & k_2 \end{bmatrix} \\ &= \begin{bmatrix} k_1 & -k_1 & 0 \\ -k_1 & k_{1+}k_2 & -k_2 \\ 0 & -k_2 & k_2 \end{bmatrix} \end{aligned} \quad (9)$$

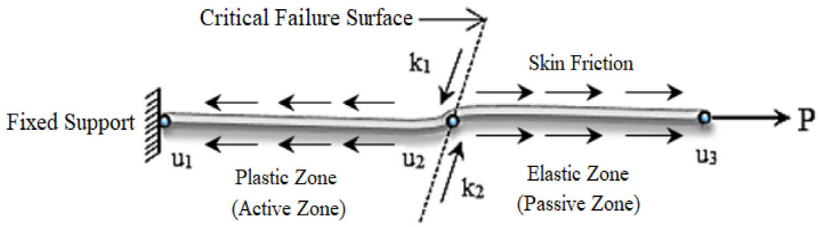
In matrix form, the force–displacement equation is

$$\begin{bmatrix} k_1 & -k_1 & 0 \\ -k_1 & k_{1+}k_2 & -k_2 \\ 0 & -k_2 & k_2 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} R \\ 0 \\ P \end{Bmatrix} \quad (10)$$

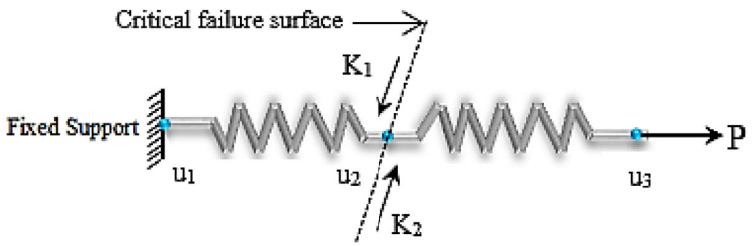
where k_1 is a force on the slip surface resisted by soil, k_2 is a force on the slip surface resisted by nail, u_1 is the displacement at the fixed support (considered as 'Zero'), u_2 is the displacement at the slip surface, and u_3 is the



(**a**) Actual model



(**b**) Activity of internal forces in soil nail



(**c**) Spring models develop from the actual model

Figure 10. Mathematical model single-degree freedom system.

Table 3. Displacement of nails observed near the slip surface in the soil mass at various nail inclinations.

Nails inclination	Forces on slip surface resist for soil (k_1) (kN/m)	Forces on slip surface resist for nails (k_2) (kN/m)	Nail 1		Nail 2		Nail 3		Nail 4		Nail 5	
			u_2 (m)	u_3 (m)	u_2 (m)	u_3 (m)	u_2 (m)	u_3 (m)	u_2 (m)	u_3 (m)	u_2 (m)	u_3 (m)
0°	380.28	141.56	0.041	0.165	0.070	0.258	0.117	0.429	0.145	0.532	0.268	0.990
5°	385.97	156.77	0.046	0.159	0.069	0.239	0.115	0.399	0.143	0.494	0.266	0.922
10°	389.25	188.49	0.046	0.141	0.069	0.213	0.116	0.354	0.143	0.439	0.267	0.819
15°	402.27	198.65	0.046	0.139	0.068.	0.205	0.113	0.340	0.139	0.422	0.260	0.785
20°	399.05	168.94	0.045	0.152	0.068	0.229	0.113	0.381	0.141	0.473	0.262	0.881
25°	397.20	160.23	0.045	0.158	0.069	0.239	0.114	0.397	0.141	0.492	0.263	0.917

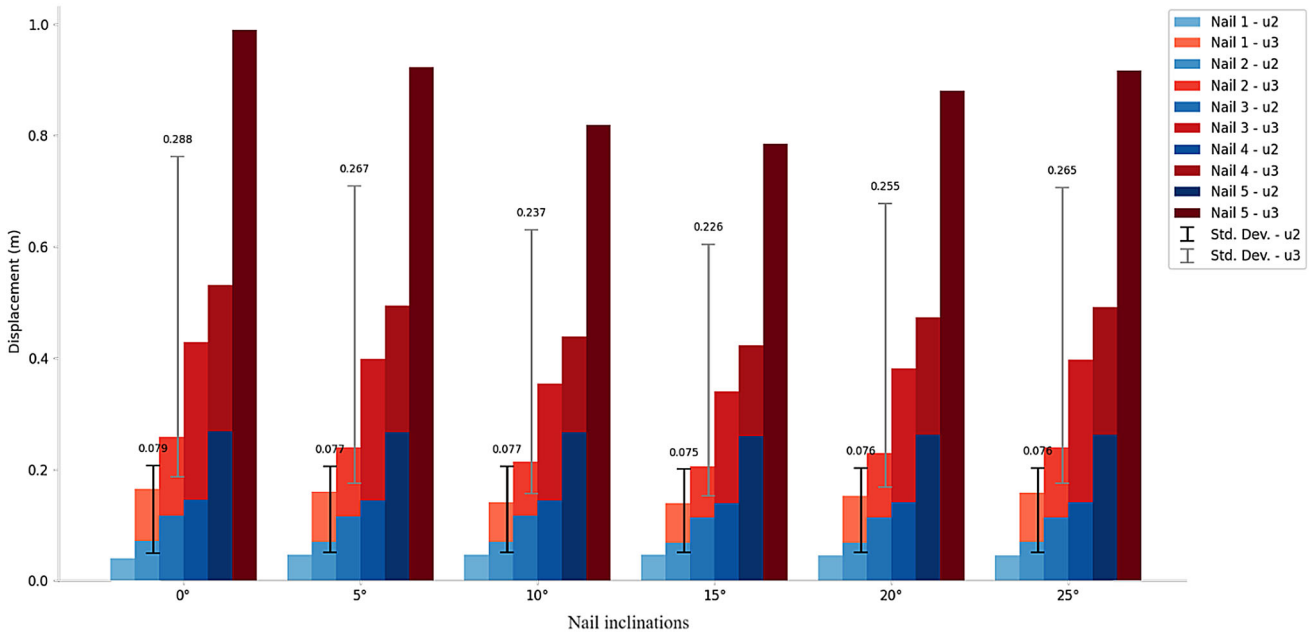


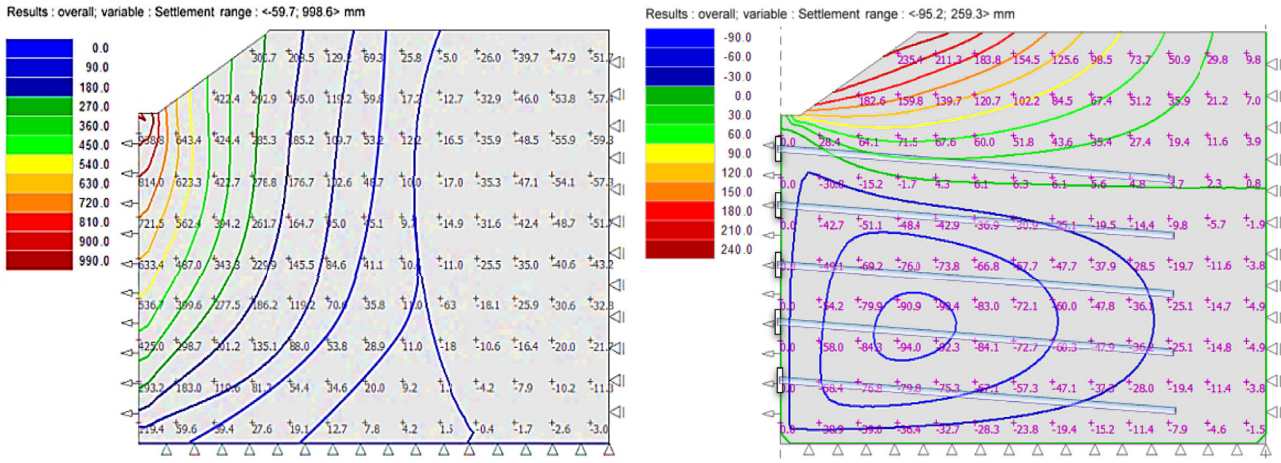
Figure 11. Standard deviation observation and correlation between displacement at node u_2 and u_3 by matrix method considering different nail inclinations.

displacement at the end of the soil nail in soil mass and P is the force acting in the nail (x-direction '+'ve).

4.1b *Boundary condition application:* The stiffness matrix will be singular if no boundary conditions are mentioned for the structural equation model. Alternatively, the determinant will be zero, and the inverse will not exist. The structure will be free to move as a rigid body and will not resist any applied load if adequate support conditions are not specified. When finite non-zero displacement values are provided, non-homogeneous boundary conditions can occur. There are two types of boundary conditions: (i) homogeneous boundary conditions and (ii) non-homogeneous boundary conditions. Homogeneous boundary conditions occur when movement is completely prohibited. While Non-homogeneous boundary conditions occur when finite non-zero displacement values are specified. By directly eliminating the rows and columns corresponding to zero displacement and degrees of freedom, the homogeneous boundary conditions are enforced in the elimination method. The displacement at node 1 = 0, for fixed. When the first row and column are removed, the equation becomes:

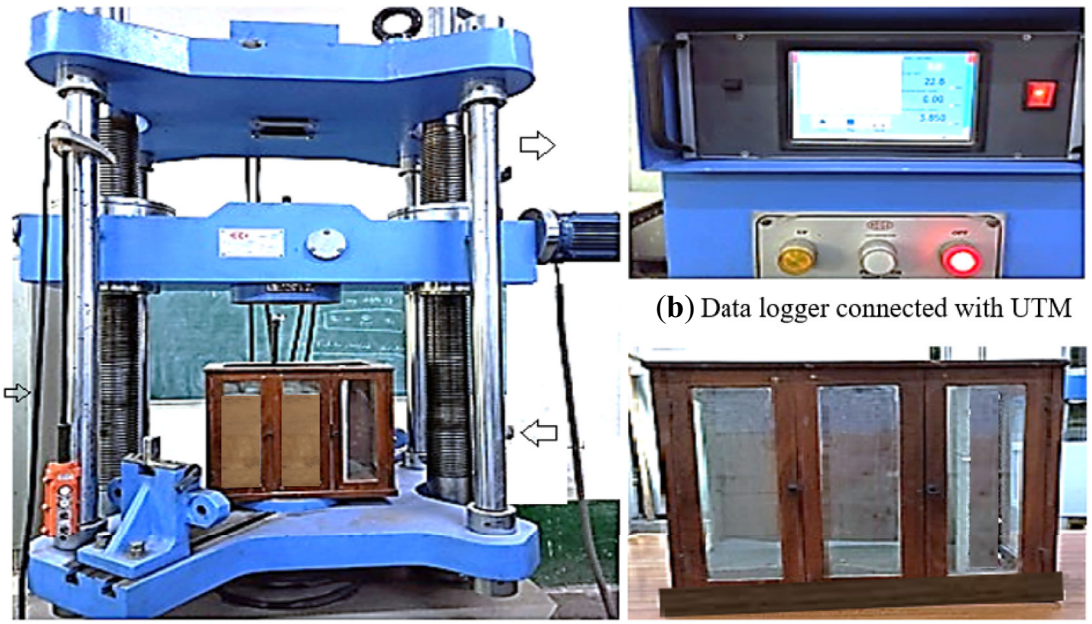
$$\begin{bmatrix} k_1+k_2 & -k_2 \\ -k_2 & k_2 \end{bmatrix} \begin{Bmatrix} u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ P \end{Bmatrix} \text{ OR } \begin{Bmatrix} u_2 \\ u_3 \end{Bmatrix} = \begin{bmatrix} k_1+k_2 & -k_2 \\ -k_2 & k_2 \end{bmatrix}^{-1} \begin{Bmatrix} 0 \\ P \end{Bmatrix} \tag{11}$$

The model here is a global stiffness matrix for determining the displacement of the nail at u_2 and u_3, by using equation (11). Where P is the force acting in the nail. Using



(**a**) Soil slope without soil nailing (**b**) Soil slope with soil nailing

Figure 12. Settlement of embankment under surcharge loading by 2D FEM



(**b**) Data logger connected with UTM

(**a**) Computerized Universal Testing Machine (UTM) (**c**) Physical Model

Figure 13. Experimental setup: soil slope model mounted on computerised UTM.

the above stiffness matrix, at the various nail inclinations, the nodal displacement is observed separately for each nail. The displacements of nails in soil mass concerning their inclination are computed and stated in table 3.

Node u_2 reports standard deviation values of 0.079, 0.077, 0.077, 0.075, 0.076, and 0.076 m for 0°, 5°, 10°, 15°, 20°, and 25°, respectively. Similarly, node u_3 reports values of 0.288, 0.267, 0.237, 0.226, 0.225, and 0.265 m for the same respective nail inclinations shown in figure11. From these observations, it appears that soil slopes at 15° nail inclinations indicate smaller displacements with an average displacement of 0.1252 m and a standard deviation of 0.075 m at node u_2. At u_3, the average displacement is 0.3782 m with a 0.226 m standard deviation. The minimum displacements at nodes u_2 and u_3 are observed at 15° nail inclinations (refer to figure11). After 15° nail inclinations,

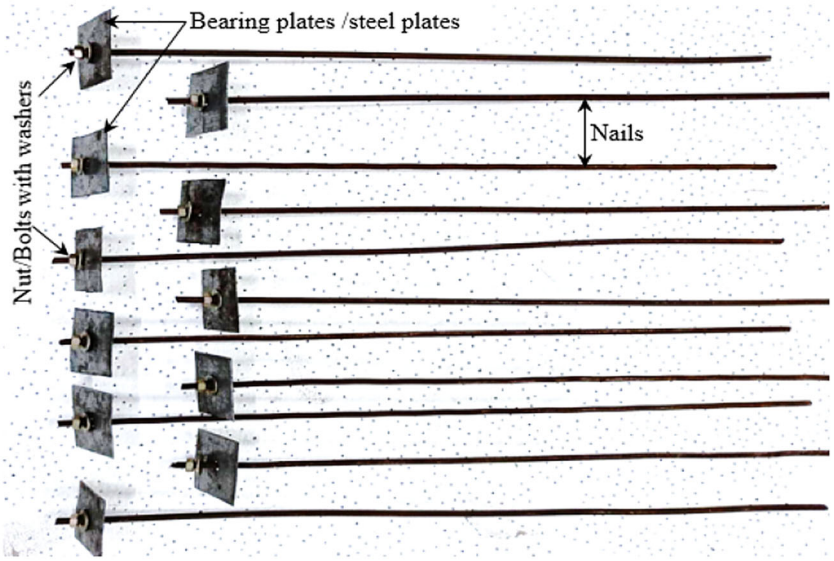


Figure 14. Nails connection/arrangement with bearing plates by using nuts/bolts.

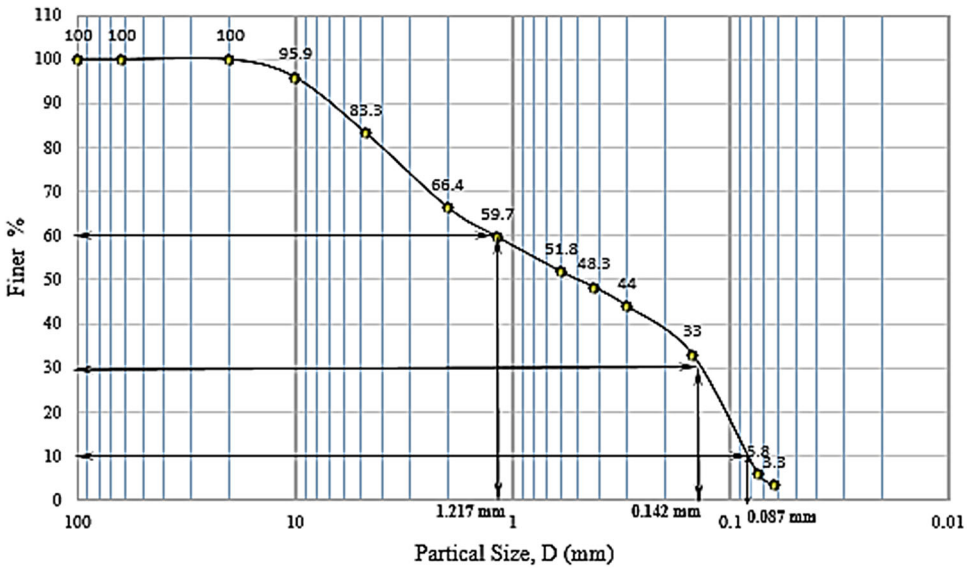


Figure 15. Grain size distribution curve for used soil sample.

displacements and standard deviations increase at nodes u_2 and u_3, indicating a reduction in slope stability.

5. Identification of settlement embankment within the soil slope by 2D FEM

Notably, it was found that a 15° nail inclination increased the FOS and overall stability of the structure, as discussed above in the context of thorough analyses and their outputs in sections 3 and 4. In this section, numerical simulations have been conducted to investigate the effects of 15° nail inclination using 2D-FEM by using Geo5 fine software. Using the properties listed in table 1 and focused on the soil slope profile shown in figure 2, a thorough 2D-FEM analysis was carried out. Both the soil structure without nail reinforcement and the soil structure reinforced with soil nails were simulated under surcharge loading conditions. The 2D-FE analysis indicated some interesting results. The structure with a nail inclination of 15° was the most stable arrangement for the slope. While the structure without soil nails showed settlement rates between <-59.7 and >998.6 mm (figure 12a), the reinforced soil slope's

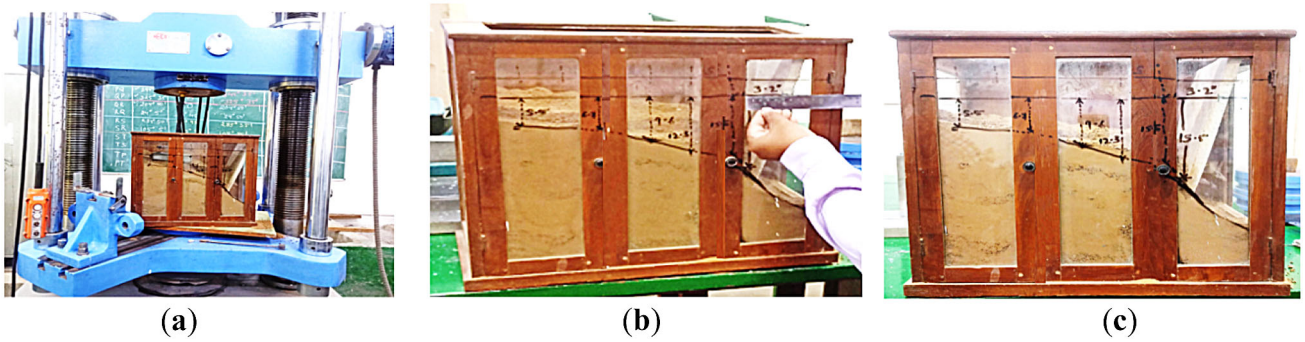


Figure 16. Observation of soil slope model-I without reinforcement: (a) Model mounted on UTM, (b) Displacement measurement, (c) Displacement within the slope

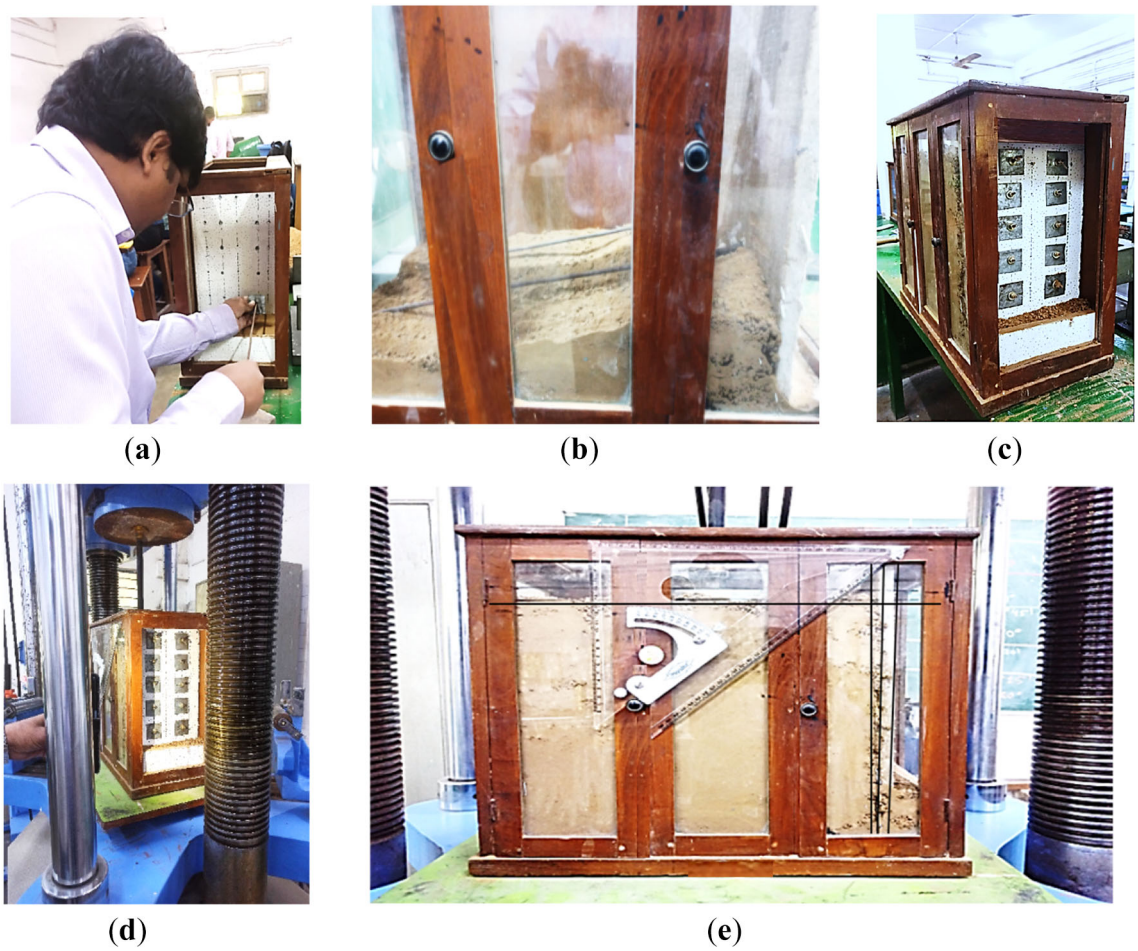


Figure 17. Complete preparation of reinforced soil nailed slope models-II from installation to performance evaluation: (a) Installation of soil nails/bars in soil slope (b) Preparation of nail inclinations (c) Complete soil nailed wall model (d) Soil nailed model mounted on UTM (e) Measurement of displacement in soil nailed wall and slope.

settlement rates varied from <-95.2 to >259.3 mm (figure 12b). Particularly, the structure without soil nails (figure 12a) experienced significantly higher settlement near the embankment face, as compared to the reinforced structure shown in figure 12(b), which experienced minimal or negligible settlement due to the presence of the attached nail within the wall.

6. Experimental investigation

In this study, the model used (figure 2) was scaled down to a physical model with dimensions of 800 mm (Length) × 250 mm (Breadth) × 400 mm (Height). This physical model was then used for experimental investigation, which was carried out using a computerized universal testing machine and a data logger. The experimental setup used for this investigation is shown in figure 13 and the arrangement of preparation of the model shown in figure 17. To analyse the reinforcement, 10 steel bars of 5 mm diameter and length of 450 mm were used, along with 40 mm × 40 mm steel plates called bearing plates, which are fixed at the facing of the wall with 8 mm diameter bolts/nuts (as shown in figure 14). For embedding the reinforcement, locally available soil from the South Delhi region is used, and its classification determined through sieve analysis. The soil is classified under Poorly Graded Sand (SP) based on IS: 1498–1970 [34]. The result of the sieve analysis is plotted in figure 15. Also, in the analytical investigations, the same type of soil classification was used.

In this experimental study, two models were represented: one is without reinforcement and the other with reinforcement (refer to figures 16 and 17). For the physical model, a 15 mm thickness wall is placed at 600 mm and filled with the 600 × 250 × 300 mm (L × B × H) soil sample (SP), with reinforcement placed at different nail inclinations (refer to figure 17b). Results showed that the reinforced model provided more strength than the one without reinforcement. During the experimental study, it was observed that the displacement of the soil-nailed wall is minimum at 15° of nail inclination with the horizontal plane. The wall displacement and slope settlement readings were taken in the horizontal and vertical directions (refer to figures 16 and 17).

Under the surcharge load, the maximum wall deflections ($\delta\Delta$) at 0°, 5°, 10°, 15°, 20°, and 25° nails inclination with the horizontal plane were observed (average) as 28, 25, 21, 16, 18, and 20 mm, respectively. The maximum wall deflection ($\delta\Delta$) without soil nail is observed as 32 mm (average) at the top of the wall. The results indicated that soil nailing is a more versatile and effective technique in slope stabilization as it provides global stability to the whole structure at 15° of nail inclination.

7. Comparative analysis between various methods used in this research

According to the findings of mathematical modelling, numerical analysis (FEM), and experimental investigation, soil nail structures are safer when soil nails are inclined at 15° with the horizontal plane. Furthermore, it was observed that soil displacement/settlement occurred under the surcharge loading. When comparing the reinforced soil slope to the unreinforced soil slope, more displacement was observed within the unreinforced slope. The experimental method finds the maximum reduction in settlement at 15° nail inclination. This is due to the reinforced soil layer's increased stiffness, which reduces the amount of settlement and lateral displacement. Furthermore, soil reinforcement increases the embankment's bearing capacity and lowers the risk of failure due to surcharge loading. At 15° nail inclination, soil reinforcement reduces settlement by 61.90 % (experimental model) and 62.71 % (finite element analysis). The experimental method result is close to the findings of the FEM. In terms of settlement reduction at 15° nail inclination, the FEM is the most effective, with a reduction of 62.71 %. In addition, the mathematical modelling shows that the 15° nail inclination causes the least displacement at nodes u_2 and u_3 when compared to other nail inclinations (refer to section 4). Therefore at 15° nail inclination, soil reinforcement is a viable option for reducing the risk of failure under surcharge loading and providing significant reductions in settlement and lateral displacement.

8. Limitations

Despite rigorous efforts, several factors pose challenges to the precision of the results obtained in this research. The intricacies inherent in modelling soil behaviour across numerical simulations, FEM, and laboratory experiments contribute to slight variations in outcomes. Modelling accurate boundary conditions, encompassing slope geometry and loading conditions, remains a persistent challenge, introducing uncertainties that potentially affect result accuracy. Additionally, the scale effect between small-scale laboratory experiments and real-world scenarios may limit the direct translation of findings. While 2D-FEM demonstrated superior performance in this study, a comprehensive understanding of the reasons behind this advantage requires further investigation. Moreover, the reliance on certain assumptions in both numerical and physical modelling analyses introduces variability, impacting the robustness of the conclusions drawn. Addressing these limitations presents a path for future research to enhance the reliability and applicability of similar studies in this domain.

9. Conclusion

Extensive research has been conducted using experimental, numerical, and analytical methods to investigate the role of soil nails in enhancing structural load-bearing capacity. These observations, confirmed by LSD, 2D-FEM, and experimental approach, consistently showed that the positive impact of soil nails on reinforcing structures. The investigation reveals the following important findings:

- Horizontal loading creates higher shear stresses than vertical loading, impacting soil slope stability. Using soil nails in the slope can prevent structural failures or landslides.
- The Limit State Design method indicates a consistent rise in FOS up to 15° nail inclination, followed by a sudden decline beyond 15°, with a peak FOS of 1.630 at 15°.
- Numerical analysis (2D-FEM) indicates a significant 62.71% reduction in settlement with poorly graded sand using a 15° nail inclination, demonstrating its effectiveness.
- Experimental results confirm a substantial 58.33% reduction in wall displacement and a noticeable 61.90% decrease in embankment settlement, especially with poorly graded sand at a 15° inclination, showcasing its efficiency in minimizing structural movement and settlement.
- The 15° nail inclination stands out as the most effective, displaying superior performance in stabilizing slopes and achieving a higher global Factor of Safety across analytical, numerical, and experimental models using Poorly graded sand (SP).

This comprehensive study highlights the pivotal role of soil nailing, notably at a 15° inclination, in fortifying stability, reducing settlement, and ensuring structural integrity. The findings underscore the critical importance of employing precise slope stabilization techniques to safeguard public safety along highways and railway embankments.

Acknowledgements

The work presented in this article is encouraged by Delhi Technological University, Civil Engineering Department, and G B Pant, DSEU, Okhla-III, Campus, New Delhi. Special thanks to the co-author of this manuscript for their valuable and constructive suggestions during the planning and development of this research work.

Author contributions PCR: Methodology, validation, formal analysis, investigation, software analysis, performed the experiments, writing—original draft and review. AKS: Conceptualization, methodology, resources, review, editing, supervision.

Conflict of interest The authors declare that there are no conflicts of interest.

References

[1] Liu J, Shang K and Wu X 2016 Stability analysis and performance of soil-nailing retaining system of excavation during construction period. *J. Perform. Constr. Facil.* 30: C4014002. https://doi.org/10.1061/(asce)cf.1943-5509.0000640

[2] Pak A, Maleki J, Aghakhani N and Yousefi M 2021 Numerical investigation of stability of deep excavations supported by soil-nailing method. *Geomech. Geoeng.* 16: 434–451. https://doi.org/10.1080/17486025.2019.1680878

[3] Elahi T E, Islam M A and Islam M S 2022 Parametric assessment of soil nailing on the stability of slopes using numerical approach. *Geotechnics* 2: 615–34. https://doi.org/10.3390/geotechnics2030030

[4] Sivakumar Babu G L and Singh V P 2011 Reliability-based load and resistance factors for soil-nail walls. *Can. Geotech. J.* 48: 915–30. https://doi.org/10.1139/t11-005

[5] Carlos A, Victor E, David E and Paul J S 2003 *Manual for Design and Construction Monitoring of Soil Nail Walls*. Federal Highway Administration, Washington, D C Report, FHWA0-IF-03-017

[6] Hu H and Lin P 2020 Analysis of resistance factors for LRFD of soil nail pullout limit state using default FHWA load and resistance models. *Mar. Georesour. Geotechnol.* 38: 332–348. https://doi.org/10.1080/1064119X.2019.1571540

[7] Fan C C and Luo J H 2008 Numerical study on the optimum layout of soil-nailed slopes. *Comput. Geotech.* 35: 585–99. https://doi.org/10.1016/j.compgeo.2007.09.002

[8] Guang-Hui C, Jin-Feng Z, Yu-Ming S and Jing-Yu C 2022 Stability assessment of soil-nailed slopes using the homogenisation approach. *Eur. J. Environ. Civ. Eng.*. https://doi.org/10.1080/19648189.2022.2052969

[9] Liu H, Tang L and Lin P 2017 Maximum likelihood estimation of model uncertainty in predicting soil nail loads using default and modified FHWA simplified methods. *Math. Probl. Eng.*. https://doi.org/10.1155/2017/7901918

[10] Lin P, Bathurst R J and Liu J 2016 Statistical evaluation of the FHWA simplified method and modifications for predicting soil nail loads 2016. *J. Geotech. Geoenviron. Eng.* 143: 1–11. https://doi.org/10.1061/(ASCE)GT

[11] Sivakumar Babu G L, Rao R S and Dasaka S M 2007 Stabilisation of vertical cut supporting a retaining wall using soil nailing: a case study. *Ground Improv.* 11: 157–162. https://doi.org/10.1680/grim.2007.11.3.157

[12] Majidian S and Komak Panah A 2020 Hybrid experimental and numerical approach for assessment of non-linear dynamic behavior of soil-nailed retaining walls. *Int. J. Non Linear Mech.* 123: 103476. https://doi.org/10.1016/j.ijnonlinmec.2020.103476

[13] Mitchell J K 1981 Soil improvement: state-of-the-art, international society for soil mechanics and geotechnical. *Engineering* 4: 509–65

[14] Munfakh G A and Wyllie D C 2000 Ground improvement engineering-issues and selection. In: *ISRM International Symposium*, Melbourne, Australia, November 2000, ISRM-IS-2000-008

[15] Jewell R A and Pedley M J 1992 Analysis for soil reinforcement with bending stiffness. *J. Geotech. Eng. ASCE* 118: 1505–28

[16] Gunawan I, Surjandari N S and Purwana Y M 2017 The study on length and diameter ratio of nail as preliminary design for slope stabilization. *J. Phys. Conf. Ser.* 909: 012073. https://doi.org/10.1088/1742-6596/909/1/012073

[17] Hu Y and Lin P 2018 Probabilistic prediction of maximum tensile loads in soil nails. *Adv. Civ. Eng.*. https://doi.org/10.1155/2018/3410146

[18] Cheuk C Y, Ng C W W and Sun H W 2005 Numerical experiments of soil nails in loose fill slopes subjected to rainfall infiltration effects. *Comput. Geotech.* 32: 290–303. https://doi.org/10.1016/j.compgeo.2005.02.005

[19] Jelušič P and Žlender B 2013 Soil-nail wall stability analysis using ANFIS. *Acta Geotech. Slov.* 2013: 61–73

[20] Arvin M R, Ghavami E and Motamedi Azari M 2021 Optimization of nail inclination angle in soil nail walls based on a prevalent limit equilibrium method. *Indian Geotech. J.* 52: 352–71. https://doi.org/10.1007/s40098-021-00574-z

[21] Baziar M H, Ghadamgahi A and Brennan A J 2021 Centrifuge study of seismic response of soil-nailed walls supporting a footing on the ground surface. *Geotechnique*. https://doi.org/10.1680/jgeot.21.00157

[22] Sivakumar Babu G L and Jaladurgam R 2015 Rehabilitation of distressed retaining walls using soil nails. *Proceedings of the Institution of Civil Engineers: Ground Improvement*;168. https://doi.org/10.1680/grim.12.00036

[23] Sivakumar Babu G L and Singh V P 2010 Soil nails field pullout testing: Evaluation and applications. *Int. J. Geotech. Eng.* 4: 13–21. https://doi.org/10.3328/IJGE.2010.04.01.13-21

[24] Bureau of Indian Standards. IS 456 (2000) 2005 *Plain and Reinforced Concrete—Code of Practice*. Tenth Reprint April. Bureau of Indian Standard

[25] Fine Civil Engineering Software 2018 *GEO5-User's Guide*. Edition 2018. Fine Civil Engineering Software

[26] Kempfert H G and Raithel M 2015 Soil improvement and foundation systems with encased columns and reinforced bearing layers. In: *Ground Improvement Case Histories: Compaction* (eds) Chu J J and Rujikiatkamjorn C, Grouting and Geosynthetics, Elsevier, Amsterdam, pp 609–33

[27] Das B M 2007 Advanced Soil Mechanics. 3rd edn. Taylor & Francis, Milton Park

[28] Sazzad M, Atful Hie A B K M and Hossain S 2016 Stability analysis of reinforcement slope using FEM. *Int. J. Adv. Struct. Geotech. Eng.* 5: 83–88

[29] Masi E B, Segoni S and Tofani V 2021 Root reinforcement in slope stability models: a review. *Geosciences (Switzerland)* 11: 212. https://doi.org/10.3390/geosciences11050212

[30] Lin H, Xiong W and Cao P 2013 Stability of soil nailed slope using strength reduction method. *Eur. J. Environ. Civ. Eng.* 17: 872–85. https://doi.org/10.1080/19648189.2013.828658

[31] Ramteke P C and Sahu A K 2023 Enhancing soil slope stability by soil nailing: a comprehensive review. *Z. Dtsch. Ges. Geowiss. (J. Appl. Reg. Geol.)*. https://doi.org/10.1127/zdgg/2023/0403

[32] Ramteke P C and Sahu A K 2022 Ground improvement technique by soil nailing: a theoretical analysis. In: *International conference on smart materials and structures, ICSMS-2022*, vol. 2810, AIP Publishing 050001–12. https://doi.org/10.1063/5.0146861

[33] Lazarte C A, Robinson H, Gomez J E, Baxter A, Cadden A and Berg R 2015 *Geotechnical Engineering Circular No. 7 Soil Nail Walls—Reference Manual*. Rep No FHWA-NHI-14-007. Washington, DC

[34] Bureau of Indian Standards. IS: 1498-1970 2007 *Classification and Identification of Soils for General Engineering Purposes*. New Delhi

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

Trans Indian Inst Met
https://doi.org/10.1007/s12666-024-03261-8

ORIGINAL ARTICLE

Tribological Behaviour of Austenitic Stainless Steel-Clad Surface Over Low Carbon Steel Produced by Cold Metal Transfer Welding Process

Varsha Mishra[1] · **N. Yuvaraj**[1] · **Vipin**[1]

Received: 23 August 2023 / Accepted: 4 January 2024
© The Indian Institute of Metals - IIM 2024

Abstract Low carbon steel has become an in-demand material in various industrial applications due to its versatility and cost-effectiveness. Despite its unparalleled advantages, it cannot be utilized in wear-resistant applications without surface alteration, which can be achieved by cladding stainless steel over low carbon steel. This study aims to clad 308L stainless steel over a low carbon steel plate using the cold metal transfer (CMT) welding process at different welding speeds (3, 4, 5, and 6 mm/sec). A ball-on-disc reciprocating tribometer was used to examine the wear characteristics of the cladding surface by varying the normal loads (30N, 40N and 50N) and frequencies (5 Hz, 10 Hz and 15 Hz). The wear resistance of the cladding surface is enhanced by 30–40% compared to a base material. Results show that the wear rate varies with the applied load and frequency. Due to the metastability of austenitic stainless steel during plastic deformation, the austenite changes into martensite. The cladding and worn-out surfaces were examined through optical and field emission scanning electron microscopy (FESEM) to study the wear mechanisms.

Keywords Cold metal transfer (CMT) · Stainless steel · Cladding · Microhardness · Wear

✉ Varsha Mishra
vrshchaubey51@gmail.com

N. Yuvaraj
yuvraj@dce.ac.in

Vipin
vipin@dce.ac.in

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

1 Introduction

Wear resistance is one of the most significant aspects of automotive and other engineering components working in wear and thermal environments [1]. Wear behaviour is the removal and transfer of surface material from inter-contact material's chemical and physical reactions under various interacting forces [2]. The different surface coating processes, such as cladding, thermal spraying, galvanizing, plating, and hard-facing, add material to the surface layer, improving the surface qualities and extending the components' life [3–6]. The cladding technique is desirable for producing sufficient thickness (more than 2.5 mm), high adhesion, simplicity, and affordability [7–10]. Components with cladding are better capable of performing any easy task in an adversarial environment for a longer period. There are several cladding processes, among which weld cladding is more suitable for economically producing thick layers [11]. CMT is an improved version of the gas metal arc welding process that produces a thick cladding surface with minimal heat input and process control over the parent metal [12]. Austenitic stainless steel is used in various wear resistance and corrosion resistance applications such as automobile brakes, heat exchangers, boiler walls in thermal power plants, gas turbines, food processing industries, nuclear reactors, biomedical implants, and chemical sectors, due to its excellent mechanical and wear resistance properties [13–18].

Rajeev et al. [19] compared the cladding of Stellite 21 over steel (H13) by CMT and standard plasma-transferred arc welding (PTAW) processes. They studied the wear characteristics of the cladding surface using a ball-on-disc tribometer. They discovered that the wear volumes, when the temperature is high, were lower than those at room temperature due to an oxide layer between the mating surfaces. Due to the larger volume % of the ε-hcp, which may be attributed

Published online: 19 February 2024

to reduced dilution and a finer microstructure, CMT coating demonstrated superior wear performance than PTAW coating. Evangeline [20] studied the wear behaviour of Inconel 625 metal deposition on 316L stainless steel using CMT and found that when a 90N force is applied, the wear rate is doubled compared to 50N and 20N loads, and the wear mechanism varies depending on the applied load. Guo et al. [21] investigated the wear behaviour of NiCrBSi/WC-Ni composite coatings over stainless steel using laser cladding. The formation of the hard WC phase and the partial dissolution of the WC particle at the Ni matrix has significantly increased the wear resistance. Pereira et al. [22] evaluated the wear behaviour of NiCoCrAlY and CoNiCrAlY laser cladding over 304 SS at low and high temperatures, and it was revealed that due to the contribution of the oxidized particles, abrasion is the primary wear mechanism at high temperatures. At low temperatures, the wear mechanism is abrasion, with some adhesion and coating delamination. The NiCoCrAlY laser cladding at high temperatures exhibits improved frictional wear behaviour, which lowers the wear rate. Peruzzo et al. [23] investigated the wear behaviour of sintered SS 316L samples after adding 0.6% and 0.8% of boron particles. A boron addition produces intergranular borides, responsible for the hardening of sintered SS 316L, lowering the coefficient of friction and preventing plastic deformation during extended sliding cycles. Davanageri et al. [24] examined the wear characteristics of heat-treated super duplex stainless steel (AISI 2507) under dry sliding conditions. The applied load was the most prominent factor, followed by sliding distance and velocity. The wear rate has a linear connection with all parameters. Jiang et al. [25] examined the wear behaviour of the diamond-like carbon coating deposited over tool steel M42 utilizing combined closed field unbalanced magnetron sputter ion plating and plasma-assisted chemical vapour deposition techniques. Soft material was predicted to display a prolonged running-in duration for a specific contact material set. Shortening the running-in period allows a material with a relatively low elastic modulus and high hardness or strength to minimize wear. It was demonstrated that sliding caused the coating's wear rate to reduce progressively to a low rate. After prolonged sliding, the wear rate tends to rise once again gradually. Rajeev et al. [26] studied the effect of reciprocating velocity, sliding distance, applied load, and counter surface temperature on Al–Si-SiCp composites. Results reveal that the load significantly impacts the coefficient of friction, and the sliding distance impacts the weight loss. The counter surface temperature had only a minimal effect on the wear. Koli et al. [27] investigated the wear behaviour of the wire arc additive processed in 308L SS samples. The sample produced with lower heat input was shown to have a low COF and wear rate due to finer grains, which enhance hardness. Also, during plastic deformation, austenite turns into martensite because of the metastability of austenite stainless steel. Nuruzzaman et al. [28] investigate the wear behaviour of steel disc materials (i.e. mild steel, 304L, and 316L) under various normal loads and sliding velocities. The results reveal that as normal load increases and sliding velocity decreases, the coefficient of friction decreases due to increased surface roughening and wear debris, which increase friction and counterface material adhesion. It was also determined that the wear rates of mild steel are much higher than those of stainless steel 304 and 316. Straffelini et al. [29] studied the sliding wear behaviour of 304L and 2205 stainless steel in dry conditions. They revealed that a large amount of austenite in 304L transformed into martensite during sliding. Plastic deformation occurs to a greater depth in 304L, which is more ductile than 2205 stainless steel. Marshall [30] and Song et al. [13] investigated the microstructure and mechanical properties of austenitic stainless steel, discovered the formation of martensite in austenitic stainless steel, and proposed that metastability in austenitic stainless steel was caused by a strain-induced transformation rather than a diffusion transformation. This would suggest that martensite is formed due to the plastic deformation of austenite. Various researchers have studied the wear behaviour of austenitic stainless steel cladding surfaces fabricated by different techniques [16, 31, 32]; also, little information exists about the CMT cladding of different alloys over base plates [19, 20]. There is a lack of study on the impact of the dry wear test variables in combination with the ball-on-disc reciprocating tribometer on 308L stainless steel cladding using the CMT process. Therefore, in the present research, considering the capabilities of CMT welding, the effect of various CMT welding process parameters on wear behaviour, as well as the microstructure characterization of the austenitic stainless steel-clad surface under dry sliding conditions, is analysed using a reciprocating tribometer (i.e. ball-on-disc), which is suitable for analysing the tribo-behaviour of the components that are having reciprocating motion, like piston ring in the cylinder, camshaft, etc. Ball-on-disc tribometers are more accurate for mechanical systems with high-load contacts and high contact pressures.

2 Materials and Methods

The low carbon steel base material and ER 308L filler wire (1.2 mm diameter) chemical compositions are exhibited in Table 1. The CMT (TPS400i) welding machine experimental setup, as shown in Fig. 1a, was used to generate the cladding surface over the base plate. The shielding gas of argon (99.99%) was used at a constant flow rate of 15 l/min. Experiments were conducted at 3 mm/sec, 4 mm/sec, 5 mm/sec, and 6 mm/sec welding speeds. After several preliminary trials, a constant welding current of 175A

Table 1 Base plate and filler wire chemical composition (% weight)

	C	Si	Mn	P	S	Ni	Cr	Al	Mo	Fe
Low carbon steel	0.13	0.16	0.58	0.02	0.02	0.02	0.04	0.02	–	Balance
ER 308L	0.02	0.42	1.42	0.02	0.012	10.03	19.15	–	0.25	Balance

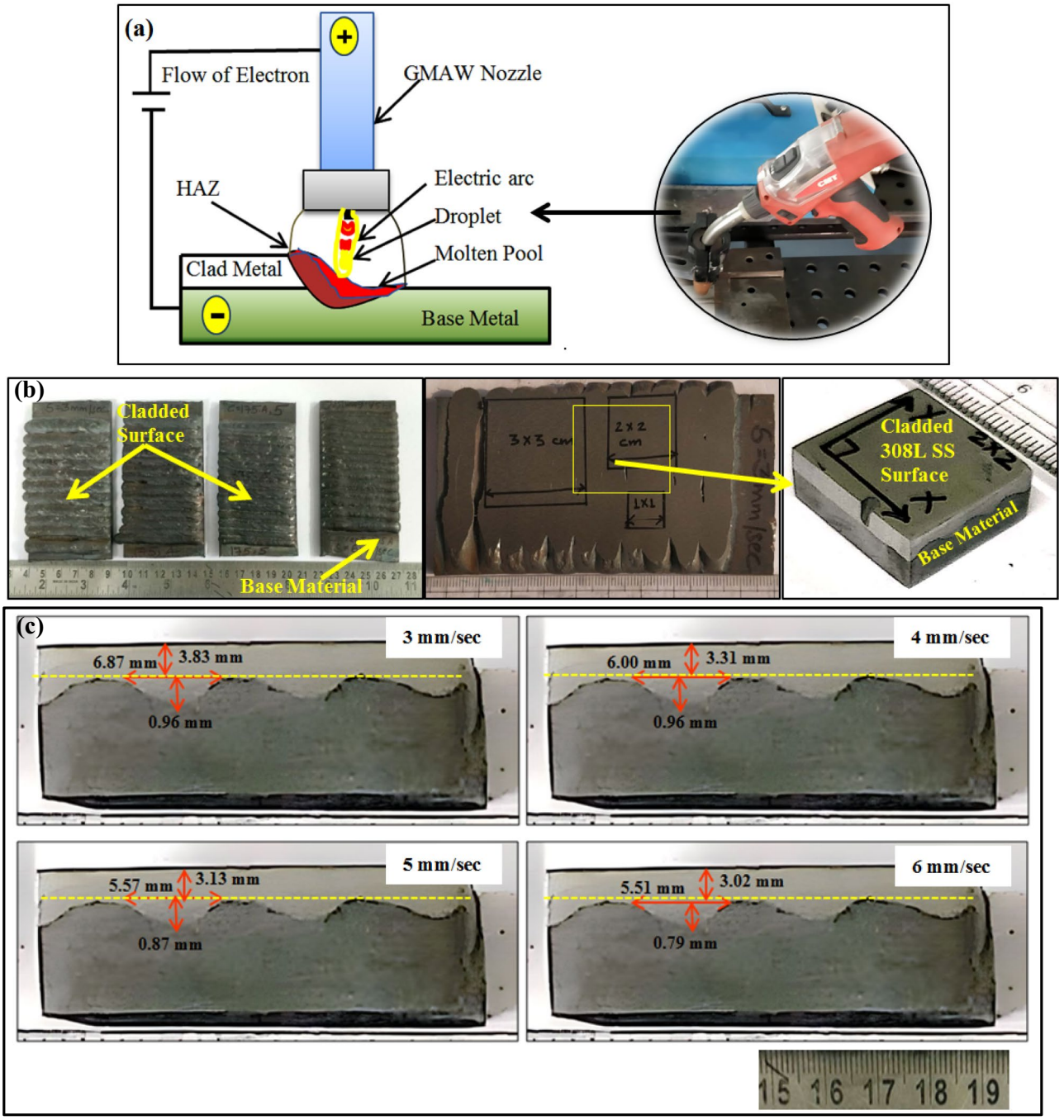


Fig. 1 **a** CMT experimental setup **b** 308L SS cladding on low carbon steel plate and extracted sample for wear test **c** cross section of clad sample extracted at different welding speeds

and a contact tip to work distance of 10 mm were fixed. Figure 1b shows the 308L SS cladding on the low carbon steel plate and the specimen extracted from the cladding surface, and Fig. 1c shows the cross section of the clad

sample extracted at different welding speeds to examine the effect of welding speed on clad surface characteristics and its wear behaviour.

The standard metallographic procedure was followed to prepare the cross section of the cladding specimen for microstructural studies. The polished samples were etched with HCl (30 ml) + distilled water (120 ml) + $FeCl^3$ (10 g) [33]. The microstructural analysis was done using an optical and field emission scanning electron microscope (FESEM). The microhardness of the cross-sectional samples was evaluated using a Vickers microhardness tester at a load of 500 g and a dwell time of 10 s.

Wear tests were conducted on the ball-on-disc reciprocating tribometer (DUCOM) as per the ASTM G133-05 standard, as shown in Fig. 2a and b. Various parameters, such as load, frequency (sliding velocity), stroke, and sliding distance, affect the wear properties. The load and frequency parameters were selected for wear studies based on the literature review and number of trial runs. In the wear test, the clad specimen was kept in contact with a 6 mm diameter tungsten ball undergoing linear reciprocating motion for 15 min, by varying different loads (30N, 40N and 50N) and frequencies (5 Hz, 10 Hz, and 15 Hz) at a specific stroke length of 10 mm in unlubricated wear conditions. Before the wear test, the specimens were grounded on a 1500-grade emery sheet and cleaned with acetone. Before and after the wear test, the wear samples were weighed. FESEM with EDS analysis was used to study the wear mechanisms. The wear surface and clad surface phase constitutions were determined with an X-ray diffractometer (XRD).

3 Result and Discussions

3.1 Microstructural Analysis

Figure 3a–e shows the microstructures of the fusion region and the clad bead regions at different welding speeds. The base metal consisted of refined and equiaxed austenite grains. The dendritic arms grow along the direction of the heat flow. The clad bead comprises a few columnar dendrites that are growing epitaxially. Columnar dendrites develop parallel to the direction of deposition. Cellular dendrites have been missing in the contact area, as shown in Fig. 3a. A deeper examination reveals that the delta ferrite has developed across most of the fusion border region, as Fig. 3b shows. This melting and re-solidification were caused by the continuous boundary layer, which forms when a small quantity of the nearby base metal melts in situ due to the superheat released from the weld pool [34]. Figure 3c shows the presence of cellular dendritic austenite that contains δ –ferrite in the clad bead region. Cellular dendrites without secondary arms, austenite grains, and equiaxed grains characterize the microstructure morphology from the substrate to the clad area due to the increasing rate of grain growth and decreasing temperature gradient. The molten pool's

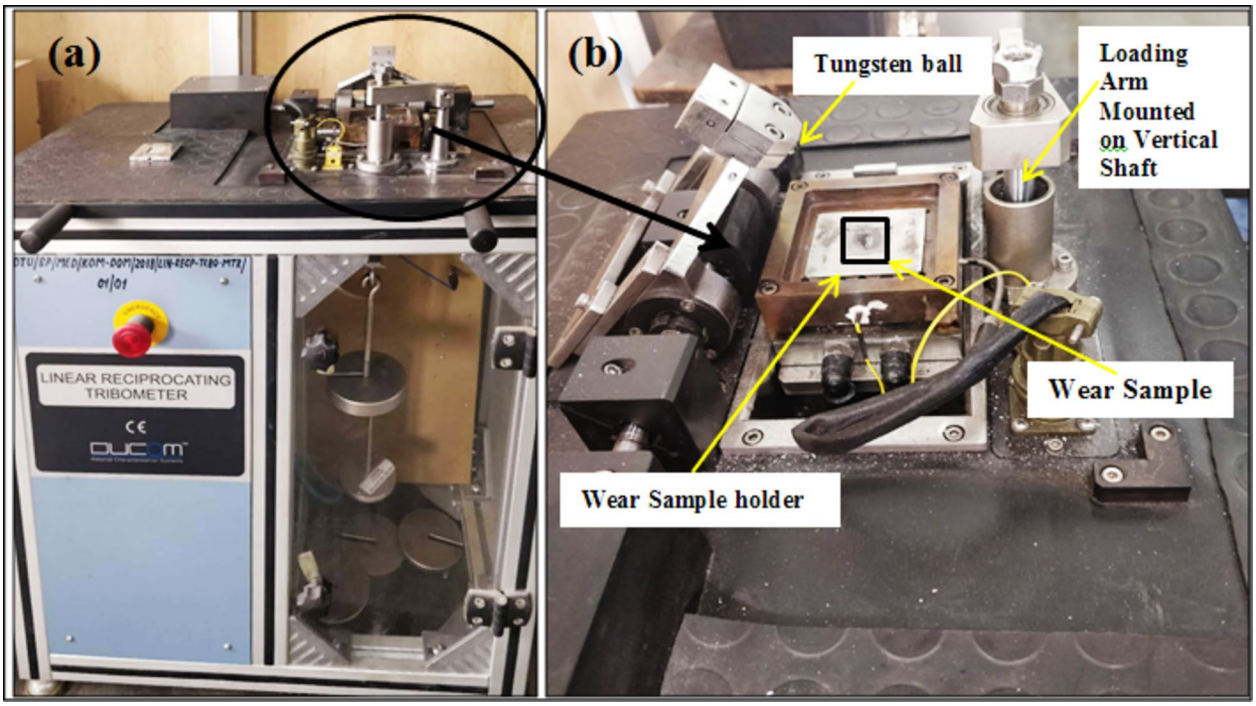


Fig. 2 **a** Linear reciprocating tribometer **b** experimental setup for the wear test

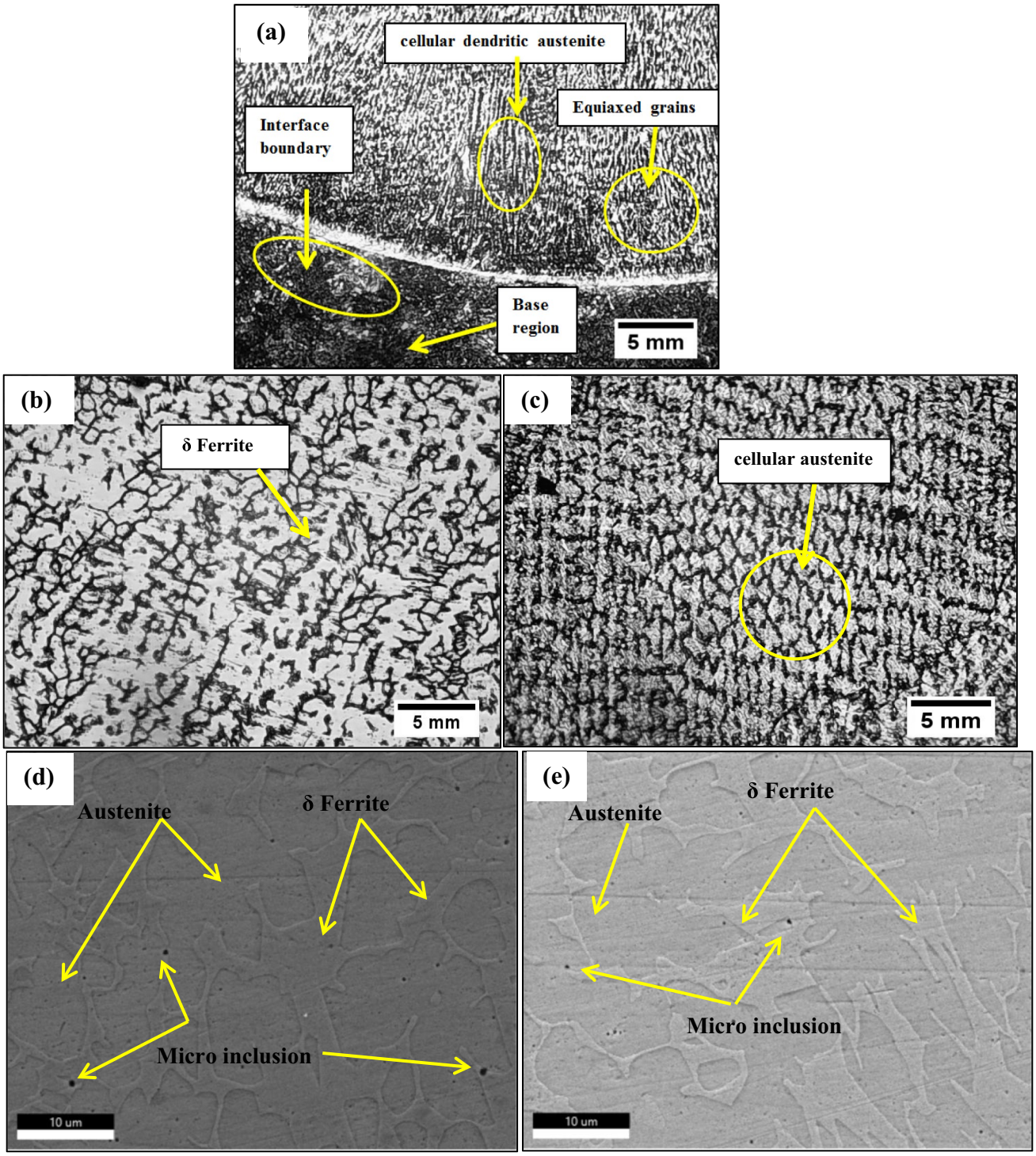


Fig. 3 Microstructure of **a** clad sample **b** Fusion region of the clad sample **c** cellular austenite in the clad region **d** FESEM image of the clad surface at 3 mm/sec welding speed **e** FESEM image of clad surface at 6 mm/sec welding speed

temperature gradients mainly cause this growth transition, and the grain growth rate of austenite in the deposited area to the right of the cladding line exhibits no evidence of retained ferrite. Figure 3d and e illustrate the FESEM image of a clad surface prepared at 3 mm/sec and 6 mm/sec welding speeds, respectively. It is inferred from the observation that the clad surface prepared at low welding speed has more inclusions than the surface prepared at high speed due to the high heat

input, which deteriorates the surface profile of the sample and its mechanical properties.

3.2 Microhardness

The average microhardness value taken at the fusion region of different clad surface samples is exhibited in Table 2. It has been found that the heat input decreases, and the microhardness increases with increasing welding speed. This increase in microhardness with increased welding speed is due to reduced heat input, which facilitates fast cooling of the weld metal containing finer grains and reduces the delta ferrite in the austenite matrix of weld metals [35].

Figure 4 displays the microhardness of the clad specimen's cross section taken at various regions. The cladding region is found to be harder than the base material due to the presence of Cr and Ni elements transmitted from the filler material. Compared to the fusion border zone, a little reduction in hardness is seen in the fusion region because of the existence of partially unmelted grains in the fusion border zone, which the new precipitating phase of the weld metal partially adopts as nuclei during the solidification stage [35]; also due to the work-hardening effect, the weld material loses some of its original strength in the fusion zone during solidification. The hard phases found in the 308L SS clad bead operate as a protective barrier during wear testing, which increases the low carbon steel (base plate's) wear resistance.

3.3 Wear Analysis of Clad Surface

The schematic diagram of reciprocating sliding wear is shown in Fig. 5. The wear particles in the tribo-pair (Tungsten/308L SS clad surface) might react differently. During the wear test, the various wear mechanisms such as oxidation, adhesive, abrasive, and fatigue wear were adopted. The worn-out particles can stick to the counter surface through adhesive contact and react with the nearby medium, forming oxidation wear. The worn-out particles may be snagged on the surfaces and roll like a third body between two sliding surfaces, creating abrasive wear. The

Table 2 Microhardness and wear test results

S.No	Clad samples at different welding speeds (mm/sec)	Heat Input (J/mm)	Microhardness ($HV_{0.5}$)	Wear rate (mm^3/m)	Coefficient of friction
1	3	630	233 ± 15	30.36	0.7
2	4	472.5	249.85 ± 15	26.56	0.6
3	5	378	272 ± 15	23.5	0.5
4	6	315	363.24 ± 15	21.66	0.4
5	Base plate	–	175 ± 15	53.24	0.8

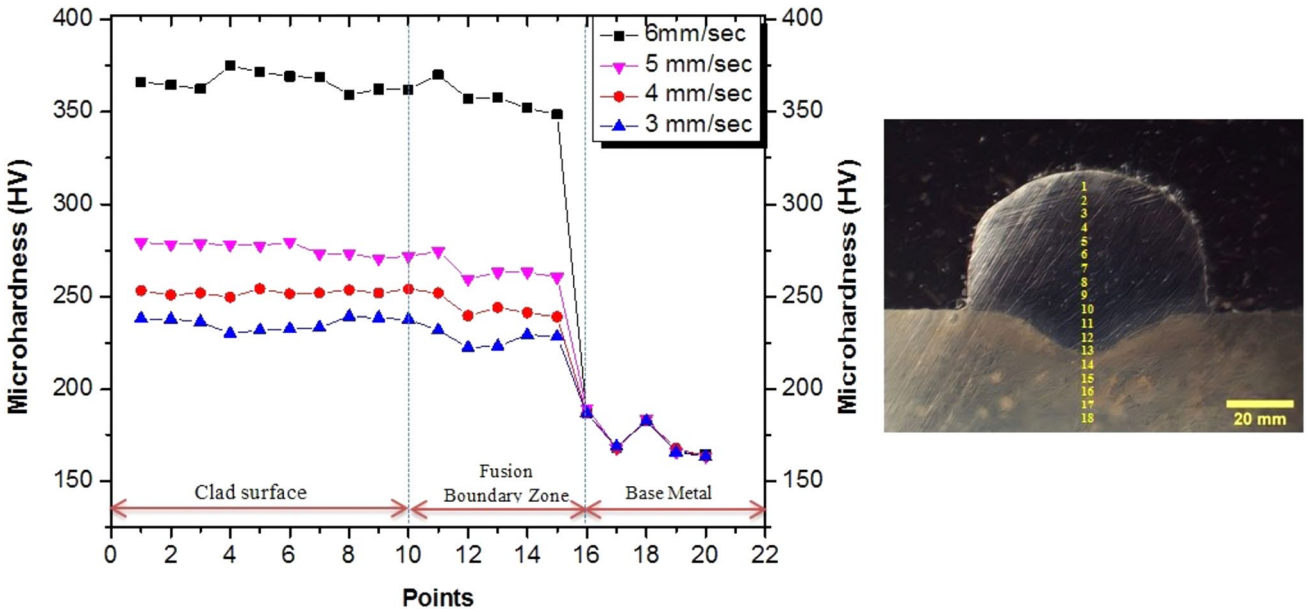


Fig. 4 Microhardness value of clad specimen taken at different regions at different welding speeds

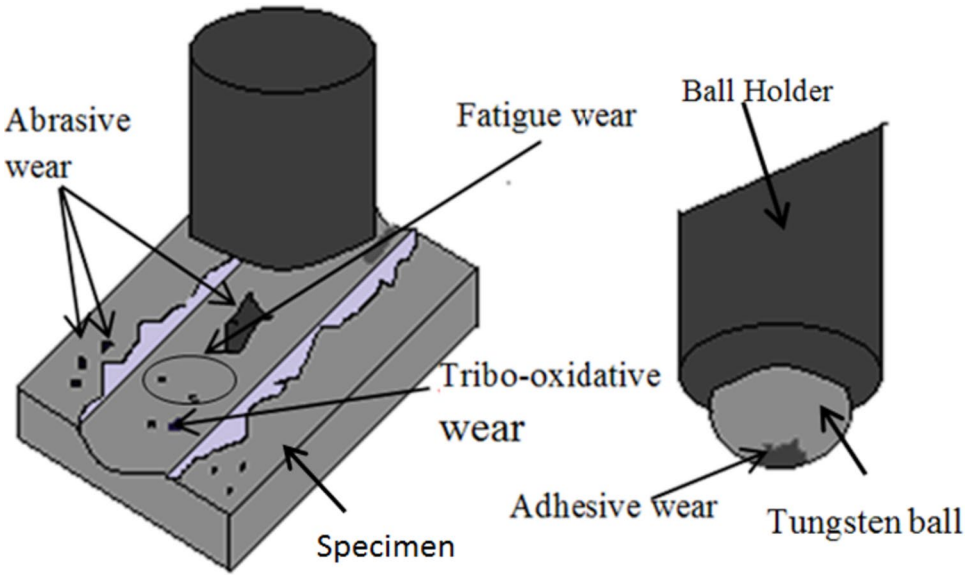


Fig. 5 Wear on the clad surface during the wear test

hard particles moving over the softer surface will form grooves on the clad surfaces. Pitting on the surface might emerge when the speed is increased considerably, i.e. fatigue wear [36, 37].

The base material and clad material wear tracks are shown in Fig. 6a and c, respectively. The base material has a broader wear track than the cladding surface. The base material's worn surface comprises pits, micro-cracks, spallation, and ploughing, as seen in Fig. 6b. Abrasion and delamination have developed during wear on low carbon steel surfaces because normal and tangential loads are transferred through the inescapably present asperities on a surface when two surfaces are in sliding contact. The subsurface deteriorates or wears away during this sliding process, with the highest plastic strain occurring beneath the contact zone. Cracks form and spread (delamination fractures) when the plastic strain increases and surpasses the critical strain for fracture (delamination), eventually losing the surface material.

On the other hand, the surface with the cladding layer has less spallation, which indicates smoother abrasion wear. The flake removal from the clad surface is primarily responsible for the emergence of fracture particles on the worn surface, as shown in Fig. 6d. Detachable abrasive particles visible while wearing them provide the impression of ploughs. The primary mechanism of wear is abrasion. Also, it alters the morphology of the worn surface and deepens the oxidized wear. This may be because the contact surface for the 308L cladding sample has a lower wear rate than the base material due to its higher hardness value.

From Table 2, it has been noticed that the wear rate of low carbon steel is reduced considerably to 30.63 (mm^3/m) from 53.24 (mm^3/m) after cladding of 308L (at 6 mm/sec welding speed). Cladding improves the COF because a steady force on a hard surface produces less contact area, resulting in a low COF and lower wear rate. The COF of the clad material decreases with increased welding speeds for constant loads and frequencies due to a reduction in heat input, which enables rapid cooling of the weld metal that contains finer grains and also reduces the delta ferrite in the austenite matrix of weld metals with increasing welding speeds. The COF and wear rate show a decreasing trend with increasing welding speed.

Also, it is found that the microhardness profiles near the wear tracks are higher than the initial microhardness value, indicating a martensitic transformation of the underlying layers during sliding, which is confirmed in the XRD plot shown in Fig. 6e. The martensite peak is absent before the wear test. However, martensite has been found in the XRD patterns of the worn-out surface because austenitic stainless steel is metastable due to a strain-induced transformation rather than a diffusion transformation, and plastic deformation converts austenite to martensite. The austenitic stainless steel martensitic transition is very prone to deformation. In austenitic stainless steel, lower sliding velocities encourage sliding wear dominated by plasticity [38]. Material transference, plastic deformation, adhesion, and mechanical mixing comprise the wear mechanism dominated by plasticity and adhesive wear, which can compete with metallic particle oxidation at extremely low sliding speeds.

3.3.1 Effect of Load on COF and Wear Rate

Figure 7a shows the typical plot of COF as a time function at different loads, and Fig. 7b shows the wear rate variations applied at different loads of the clad surface. The COF trend suddenly increases as all loads' sliding time is raised from 0 to 60 s. The COF gradually gets stable as the sliding time increases. The increase in wear rate decreases the

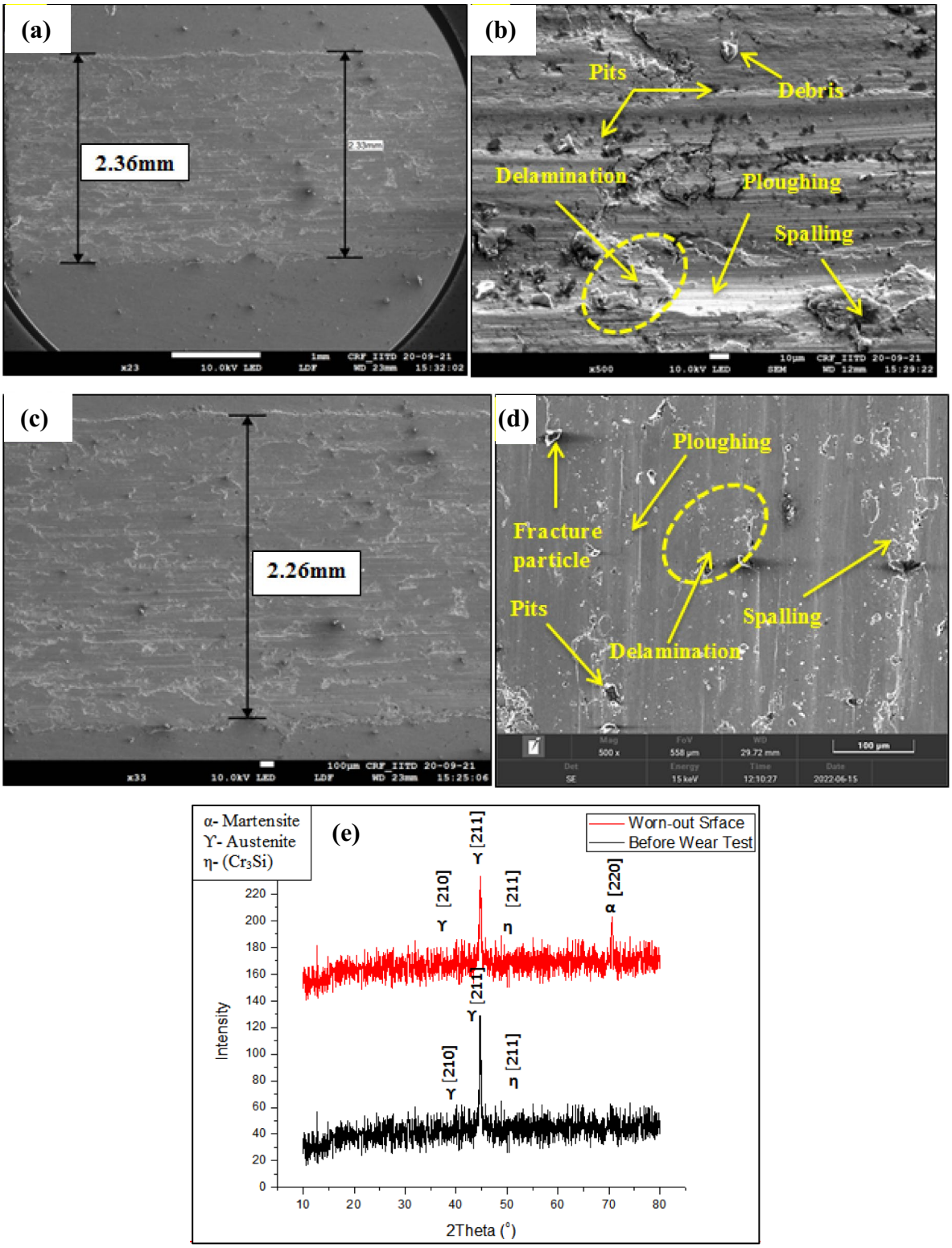


Fig. 6 FESEM micrograph of worn surface (**a**, **b**) base material (**c**, **d**) clad surface (**e**) XRD plot of wear track before and after wear test

contact area between friction pairs compared to the counterpart [39]. As the applied load increases, an increase in wear rate is observed. Also, COF increases when the normal load increases, as seen in Fig. 7a. It is stated that an increase in surface roughness and the introduction of significant amounts of wear debris account for the rise in friction

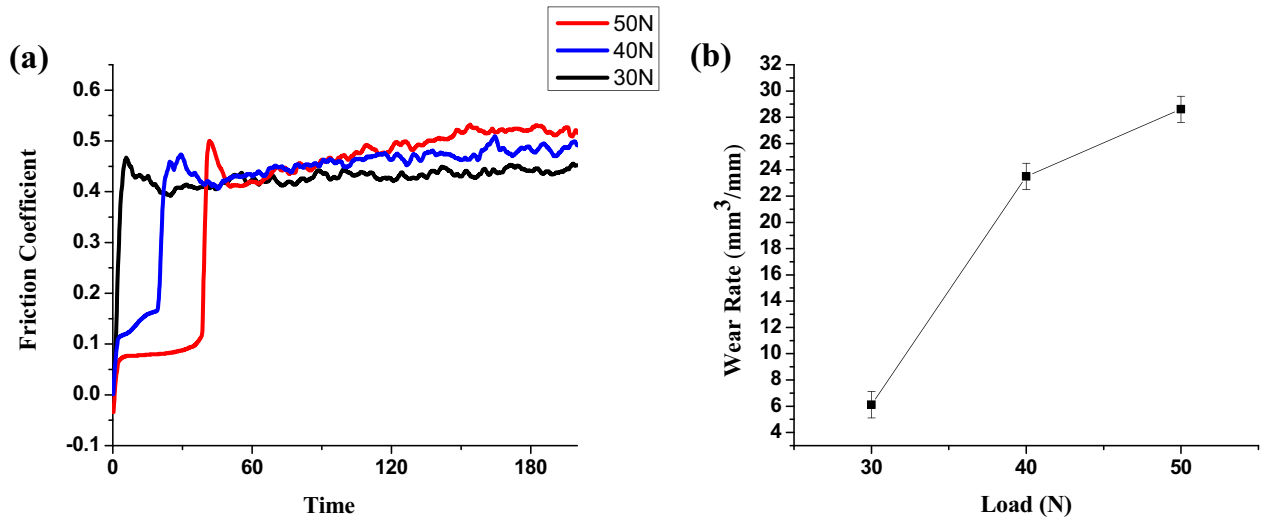


Fig. 7 **a** COF as a time function at different loads (5 Hz frequency) **b** Wear rate variation with the different loads (5 Hz frequency)

with an increase in normal load [40, 41]. Similar results are obtained for the SS 420, which is compatible with the findings of Gupta et al. [42], in that the COF increases with increasing normal load.

Figure 7b shows that the wear rate increases with increasing load. The wear rate increases to 28.61 (mm^3/m) from 6.114 (mm^3/m) on increasing load from 30 to 50N. As the load increases, the actual contact area approaches the nominal contact area, increasing the frictional force between the two surfaces. The increasing surface causes higher wear in contact and frictional force. This shows that when the applied load increases, the shear force and frictional thrust rise, increasing the wear rate. Chowdhury and Helali [43] noted similar traits of variation for low carbon steel couples, i.e. wear rate rises with a rise in normal load.

Figure 8a shows the FESEM micrographs of the wear track obtained at 40N and 5 Hz. The occurrence of tiny oxide particle agglomerates at the lowest frequency indicates that metallic particles have oxidized [44]. Due to the large exposed surface, the metallic pieces remaining at the contact can break into tiny sizes and oxidize quickly. The heat deformation speeds up oxidation, and the energy increase brought on by the surface energy and defect density increases. The surface oxidation resulting from the particles' agglomeration may shield them from further wear loss [44]. Various works have reported that austenitic stainless steels experience oxide development during sliding wear [45]. At a 40N load, the debris was created by bigger metallic particles with various morphologies, including uneven multilayered structures, plates, and spherical or cylindrical forms, as seen in Figs. 8b and c. This implies that initially separated irregular plate-shaped particles were plastically distorted and rolled to produce spherical or cylindrical debris when held within the contact. Uneven plates may crack into smaller pieces that are distorted, rolled up, or shattered [15]. Figure 8d shows the FESEM micrographs and EDS plot of the sample's wear track obtained at 50N and 5 Hz. Fewer oxides are formed at higher loads, and adhesive and abrasive wear processes are observed. Many wear particles are detached from the surface when the load and frequency increase. Due to the increased resistance of the debris to being caught, disintegrated, and oxidized to create the protective layer, the wear rate may be increased.

3.3.2 Effect of Frequency on COF and Wear Rate

Figure 9a shows the COF graph as a time function at a different frequency, and Fig. 9b shows the wear rate variation with the various frequencies. It is observed from Fig. 9a that the COF trend suddenly increases for all frequencies as the sliding time is raised from 0 to 60 s and further gradually gets stable as the sliding time increases. The COF increases with an increase in frequency (5 Hz to 15 Hz) because sliding contact between two materials causes heat to be generated at the asperities, which raises the temperature [31]. The adherence of the detachable clad surface material to the ball increases the COF with increasing frequency. The variations in wear rate with frequency are shown in Fig. 9b. Wear rates increase from 23.50 mm^3/m to 53.24 mm^3/m when the frequency increases between 5 and 15 Hz. When the frequency increases, many wear particles become detached from the surface. The wear rate increases because of the greater resistance of the particles to being captured, disintegrated, and oxidized to form the protective layer. Even if the rubbing duration is constant for all sliding speeds, the rubbing length increases with increasing frequency. The increased wear rate at high frequency is due to the larger contact area

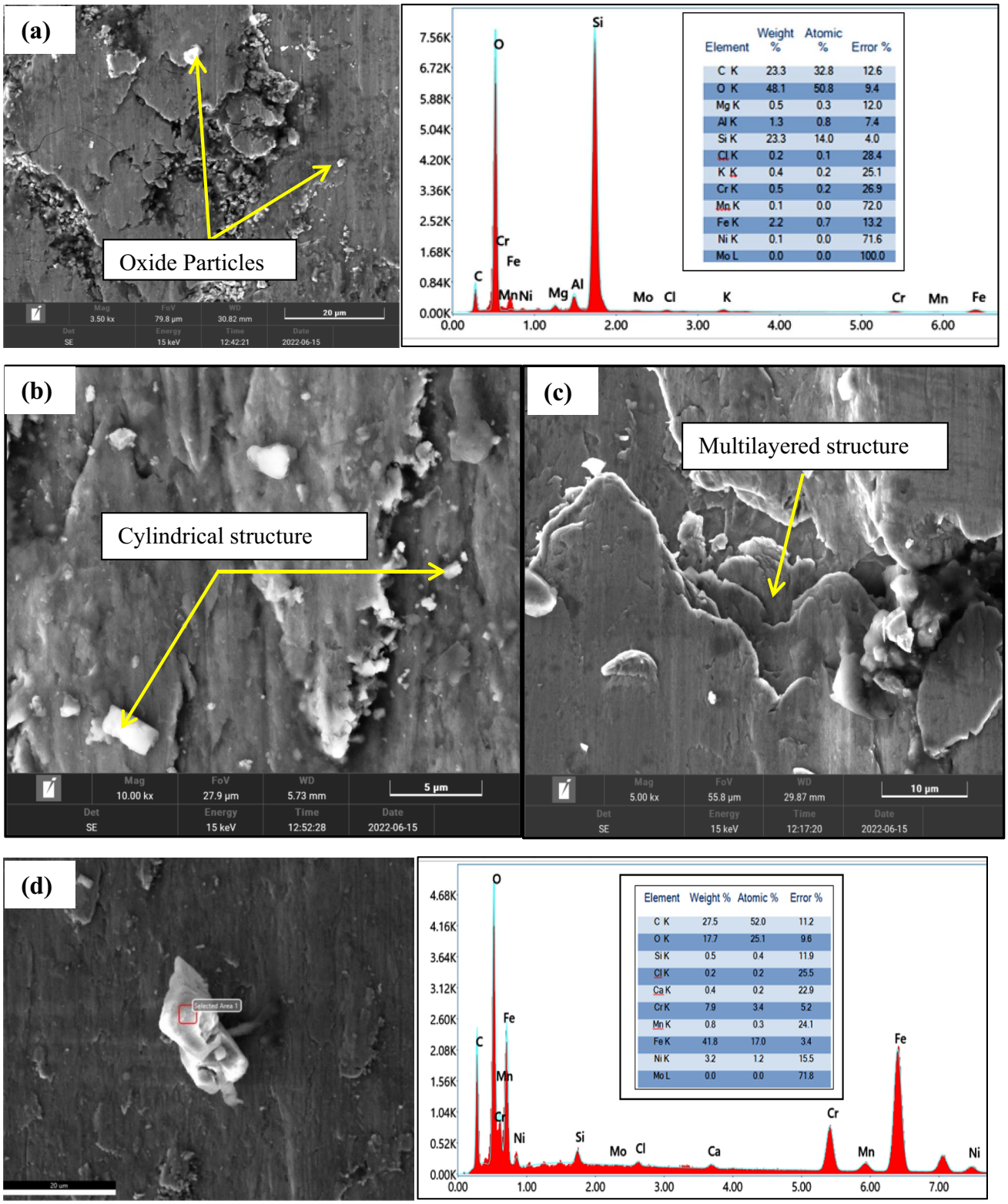


Fig. 8 **a** FESEM and EDS Plot of sample wear track obtained at 40N and 5 Hz. **b** Wear particles of clad surface exhibiting cylindrical forms (40 N load and 15 Hz frequency) **c** Clad surface revealing a multilayered structure (50 N Load 15 Hz frequency) **d** FESEM and EDS plot of wear surface at 50N and 5 Hz

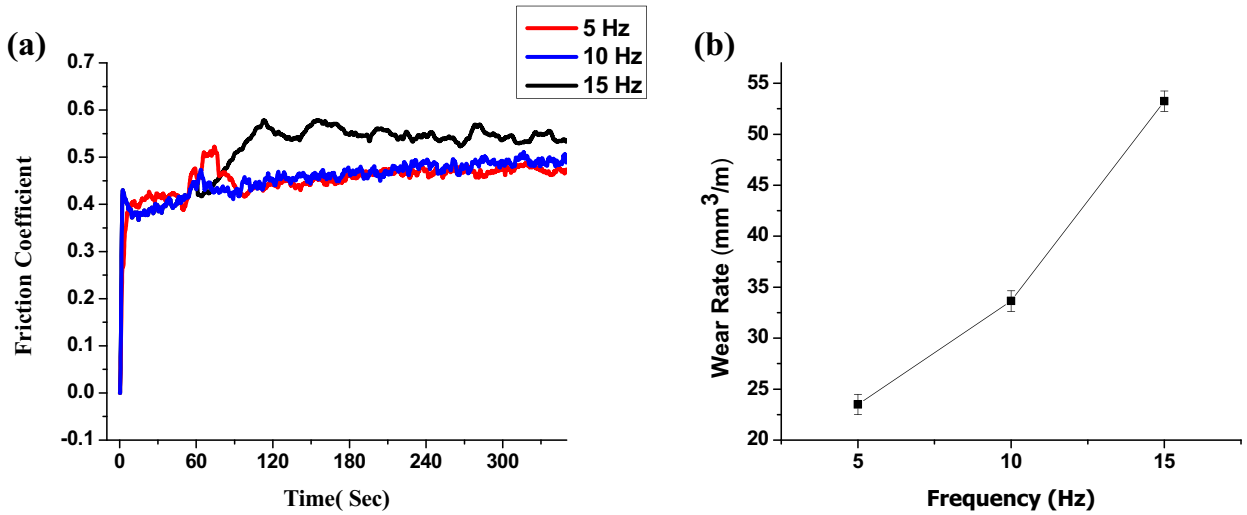


Fig. 9 **a** COF as a function of time at different frequencies (40N load) **b** wear rate variation with the variation of frequency (40N load)

between contacting surfaces and the material's lower shear strength [31].

4 Conclusions

This research used the CMT technique to examine the wear behaviour of a 308L austenitic stainless steel-clad surface. The following conclusions are drawn by evaluating the results:

1. The CMT technique is a stable process for producing defect-free clad overlay.
2. Welding speed significantly affects wear behaviour and metallurgical and mechanical properties. The low welding speed and high heat input induce low hardness, which causes more plastic deformation than the high welding speed. With an increase in welding speed from 3 mm/sec to 6 mm/sec, the wear rate decreases to 21.66 (mm^3/m) from 30.63 (mm^3/m).
3. COF values within the reported range increase as normal load and frequency increase. In the case of welding speed, COF decreases with increasing welding speed.
4. The wear rate increases from 6.114 (mm^3/m) to 28.61 (mm^3/m), increasing load from 30 to 50N. The wear rate shows a positive trend; as frequency increases, the wear rate increases.

Funding This research received no specific grant from public, commercial, or not-for-profit funding agencies.

Declaration

Competing Interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Lu P, and Wood R J, *Surf. Topogr. Metrol. Prop.* **8** (2020) 043001. https://doi.org/10.1088/2051-672X/abb6d0
2. Yin M, Yin H, and Long J, *Surf. Topogr: Metrol. Prop.* **10** (2022) 025022. https://doi.org/10.1088/2051-672X/ac5f67
3. Kirchgaßner M, Badisch E, and Franek F, *Wear* **265** (2008) 772. https://doi.org/10.1016/j.wear.2008.01.004
4. Gualco A, Svoboda H G, Surian E S, and de Vedia L A, *Mater. Des.* **31** (2010) 4165. https://doi.org/10.1016/j.matdes.2010.04.026
5. Chatterjee S, and Pal T K, *Wear* **255** (2003) 417. https://doi.org/10.1016/S0043-1648(03)00101-7
6. Mendez P F, Barnes N, Bell K, Borle S D, Gajapathi S S, Guest S D, Izadi H, Gol A K, and Wood G, *J. Manuf. Process.* **16** (2014) 4. https://doi.org/10.1016/j.jmapro.2013.06.011
7. Lee J W, Nishio K, Katoh M, Yamaguchi T, and Mishima K, *World* **49** (2005) 94. https://doi.org/10.1007/BF03266493
8. Shahi A S, and Pandey S, *J. Mater. Process. Technol.* **196** (2008) 339. https://doi.org/10.1016/j.jmatprotec.2007.05
9. Chen J H, Hsieh C C, Hua P S, Chang C M, Lin C M, Wu P T Y, and Wu W, *Met. Mater. Int.* **19** (2013) 93. https://doi.org/10.1007/s12540-013-1015-4
10. Palani P K, and Murugan N, *J. Adv. Manuf. Technol.* **30** (2006) 669. https://doi.org/10.1007/s00170-005-0101-2
11. Kiran V B, Krishna M, Natraj J R, and Kumar S, *Int. J. Eng. Technol.* **4** (2012) 18.
12. Koli Y, Yuvaraj N, and Aravindan S, *Mater. Res. Express* **6** (2020) 1265e5. https://doi.org/10.1088/2053-1591/ab61b6
13. Song R B, Xiang J Y, and Hou D P, *J. Iron Steel Res. Int.* **18** (2011) 53. https://doi.org/10.1016/S1006-706X(11)60117-9

14. Xie F, He X, Cao S, and Qu X, *J. Mater. Process. Technol.* **213** (2013) 838. https://doi.org/10.1016/j.jmatprotec.2012.12.014
15. Farias M C M, Souza R M D, Sinatora A, and Tanaka D K, *Wear* **263** (2007) 773. https://doi.org/10.1016/j.wear.2006.12.017
16. Lo K H, Shek C H, and Lai J K L, *Mater. Sci. Eng. R Rep.* **65** (2009) 39.
17. Wang Y, Yue W, Kang J, Zhu L, Fu Z, and Wang C, *J. Tribol.* **141** (2019) 042102. https://doi.org/10.1115/1.4042392
18. Kennedy F E, Ye Y, Baker I, White R R, Barry R L, Tang A Y, and Song M, *Wear* **496** (2022) 204309. https://doi.org/10.1016/j.wear.2022.204309
19. Rajeev G P, Kamaraj M, and Bakshi S R, *Surf. Coat. Technol.* **375** (2019) 383. https://doi.org/10.1016/j.surfcoat.2019.07.019
20. Evangeline A, and Sathiya P, *Mater. Res. Express* **6** (2019) 066516. https://doi.org/10.1088/2053-1591/ab0a10
21. Guo C, Chen J, Zhou J, Zhao J, Wang L, Yu Y, and Zhou H, *Surf. Coat. Technol.* **206** (2011) 2064. https://doi.org/10.1016/j.surfcoat.2011.06.005
22. Pereira J, Zambrano J, Licausi M, Tobar M, and Amigo V, *Wear* **330** (2015) 280. https://doi.org/10.1016/j.wear.2015.01.048
23. Peruzzo M, Serafini F L, Ordoñez M F C, Souza R M, and Farias M C M, *Wear* **422** (2019) 108. https://doi.org/10.1016/j.wear.2019.01.027
24. Davanageri M, Narendranath S, and Kadoli R, *Arch. Foundry Eng.* **16** (2016) 47.
25. Jiang J, and Arnell R D, *Wear* **217** (1998) 190. https://doi.org/10.1016/S0043-1648(98)00178-1
26. Rajeev V R, Dwivedi D K, and Jain S C, *Tribol Int.* **43** (2010) 1532. https://doi.org/10.1016/j.triboint.2010.02.014
27. Koli Y, Aravindan S, and Rao P V, *J. Tribol.* **145** (2023) 031706. https://doi.org/10.1115/1.4056191
28. Nuruzzaman D M, and Chowdhury M A, *Int. J. Surf. Eng. (IJSEIMS)* **1** (2013) 33. https://doi.org/10.4018/ijseims.2013010103
29. Straffelini G, Molinari A, and Trabucco D, *Metall. Mater. Trans. A* **33** (2002) 613. https://doi.org/10.1007/s11661-002-0123-4
30. Marshall P, Austenitic stainless steels: microstructure and mechanical properties, (1984).
31. Bartolomeu F, Buciumeanu M, Pinto E, Alves N, Carvalho O, Silva F S, and Miranda G, *Addit. Manuf.* **16** (2017) 81. https://doi.org/10.1016/j.addma.2017.05.007
32. Samandi M, Shedden B A, Smith D I, Collins G A, Hutchings R, and Tendys J, *Surf. Coat. Technol.* **59** (1993) 261. https://doi.org/10.1016/0257-8972(93)90094-5
33. Roy J G, and Yuvaraj N, *Trans. Ind. Ins. of Met.* **74** (2021) 2397. https://doi.org/10.1007/s12666-021-02326-2
34. Lippold J C, and Savage W F, *Weld J* **59** (1980) 48s.
35. Chuaiphan W, and Srijaroenpramong L, *J. Mater. Process. Technol.* **214** (2014) 402. https://doi.org/10.1016/j.jmatprotec.2013.09.025
36. Rigney D A, Naylor M G S, Divakar R, and Ives L K, *Mater. Sci. Eng. C* **81** (1986) 409. https://doi.org/10.1063/1.4914239
37. Selçuk B, Ipek R, and Karamış M B, *J. Mater. Process. Technol.* **141** (2003) 189. https://doi.org/10.1016/S0924-0136(02)01038-5
38. Stachowiak G W, and Batchelor A W, *Engineering tribologys*, Butterworth-heinemann (2013).
39. Rajkumar V, Arjunan T V, and Rajesh Kannan A, *J. Braz. Soc. Mech. Sci. Eng.* **42** (2020) 1. https://doi.org/10.1007/s40430-019-2092-1
40. Bhushan B, and Kulkarni A V, *Thin Solid Films* **278** (1996) 49. https://doi.org/10.1016/0040-6090(95)08138-0
41. Godse R S, Gawande S H, and Keste A A, *J. Bio Tribo. Corro.* **2** (2016) 1. https://doi.org/10.1007/s40735-016-0034-3
42. Gupta M K, Demirsöz R, Korkmaz M E, and Ross N S, *J. Tribol.* **145** (2023) 041703. https://doi.org/10.1115/1.4056423
43. Chowdhury M A, and Helali M M, *Wear* **262** (2007) 198. https://doi.org/10.1016/j.wear.2006.05.007
44. Stott F H, *Tribol Int.* **31** (1988) 61. https://doi.org/10.1016/S0301-679X(98)00008-5
45. Smith A F, *Wear* **110** (1986) 151. https://doi.org/10.1016/0043-1648(86)90143-2

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.

Validation of a Parameter Design methodology for Air-core two-coil WPT system with multi-resonant LC compensation

Akshay
Electrical Engineering Department
Delhi Technological University
Delhi,India
akshayjanagal@gmail.com

Gaurav Yadav
IEEE Graduate Student Member
Delhi Technological University
Delhi,India
gaurav812786@gmail.com

Mukhtiar Singh
IEEE Senior Member
Delhi Technological University
Delhi,India
smukhtiar_79@yahoo.co.in

2023 IEEE International Conference on Power Electronics, Smart Grid, and Renewable Energy (PESGRE) | 979-8-3503-1057-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/PESGRE58662.2023.10404443

***Abstract*—Wireless Power Transfer (WPT) possesses the revolutionary potential to replace traditional wired power transmission methods. WPT significantly enhances the usability of Electric Vehicle (EV) charging applications by facilitating the removal of obstacles within charging operations. This study demonstrates a wireless adapter that has been optimized for electric vehicle charging. This article offers a solution for optimizing the compensation network, with a particular emphasis on coil parameter specifications, its modelling, mutual inductance calculation air gap, coupling factor, and compensated basic multi-resonant LC compensation resonant network parameters calculations to reduce output voltage sensitivity and charging the EV. The proposed converter makes use of the energy from the AC grid to correct power factor while wirelessly charging the EV battery. The entire charging procedure consists of AC-DC conversion, DC-HFAC inversion, wireless power transfer, and HFAC-DC rectification. A prototype charger is constructed for a multi-resonant topology, and the MATLAB/Simulink model is used to evaluate the performance of the optimized resonant network.**

Keywords—EV Charging, Wireless Power Transfer (WPT), HFAC, Coupling Factor, Mutual Inductance, Compensation.

I. INTRODUCTION

When it comes to reducing global greenhouse gas (GHG) emissions, the transportation industry is a key offender; however, battery electric vehicles (BEVs) provide a promising answer by drastically lowering carbon emissions using renewable energy sources. [10][11]. By doing away with the need for wires,(WPT) technology facilitates this shift by revolutionising power transmission for electric vehicle charging applications and enabling fully autonomous, hands-free recharging with the proper alignment of transmitting (Tx) and receiving (Rx) coils [12],[13].

An integral part of every WPT system, high-frequency inverters improve the quality factor and power transfer efficiency [1]. Increased switching losses at higher operating frequencies are a problem for traditional hard-switched and pulse-width modulated inverters when used for WPT applications. Coil quality attributes, alignment, and coupling coefficient all have a major impact on efficiency and must be taken into account in an optimised design [4], [7].

Multi-resonant LC compensation inductive WPT systems are used in the suggested solution, together with active rectifiers on the mains side and converters on the load side, to improve the overall efficiency of the system [2], [5]. Precise control systems at the roadside ensure the best possible waveform output and automate adjustments at the vehicle end, resulting in increased efficiency [3].

The study's overarching goal is to learn more about how WPT technology is progressing of creating a future in which electric vehicle (EV) charging infrastructures are more eco-friendly and sustainable[15][16]. By delving into the nuts and bolts of WPT for EV applications, we may get a better idea of its potential effects on both the transportation industry and the natural world [6],[8], [9].

II. PROPOSED TECHNOLOGY

The wireless electric vehicle (EV) charger consists of four crucial components for efficient wireless power transmission: an AC-DC converter, a High-Frequency AC Inverter (HFAC), Wireless Power Transmission (WPT) via Inductive Power Transfer (IPT), and a Full Bridge Rectifier with Capacitive Filter as shown in Fig (1). The system transforms grid AC power into a refined DC output for an EV battery.

The utilization of Active Front End (AFE) LCL filters serves to mitigate the presence of harmonics and noise when receiving a 415V V_{PP} input power supply from a three-phase Grid. The 6-switch Bridge rectifier is utilized to transform the Grid voltage into a pulsating DC voltage of 800V. This DC voltage is subsequently stabilized by the presence of the DC-link capacitor. The control algorithm employed in the SRFT Control theory enables the voltage control loops of the AFE converter to modulate the output voltage within the ranges of V_{PP} as (700V to 900V or 800V to 600V), commonly referred to as maintaining the DC Link Voltage.

The HFAC inverter functions at a frequency of 30 kHz and employs an open-loop control system. Despite the lack of feedback, this inverter consistently produces a stable and constant output. The WPT technique known as Inductive Power Transfer (IPT) involves the utilization of magnetic induction to transfer energy from a primary coil to a secondary coil. In this procedure, the initial 800V square wave alternating current (AC) voltage undergoes conversion into an (800V) V_{PP} high frequency sinusoidal AC voltage. This resultant high frequency AC voltage subsequently generates a (1000V) V_{PP} high frequency AC voltage.

The utilization of the Full Bridge Rectifier with Capacitive Filter is employed for the purpose of converting AC power to DC power. This conversion process results in the generation of a stable output voltage that is suitable for recharging the 48V lead-acid battery commonly found in EVs. The current paper is structured as follows:

Section II examines a multi-resonant wireless power transfer (WPT) charger with a circular spiral coil with perfect alignment. LCR meter results in Section III validate the coil model. Section V result reveals validate and support the coupling factor. Section VI concludes this study by summarizing towards control of AFE which ensures UPF.

979-8-3503-1057-3/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:11:26 UTC from IEEE Xplore. Restrictions apply.

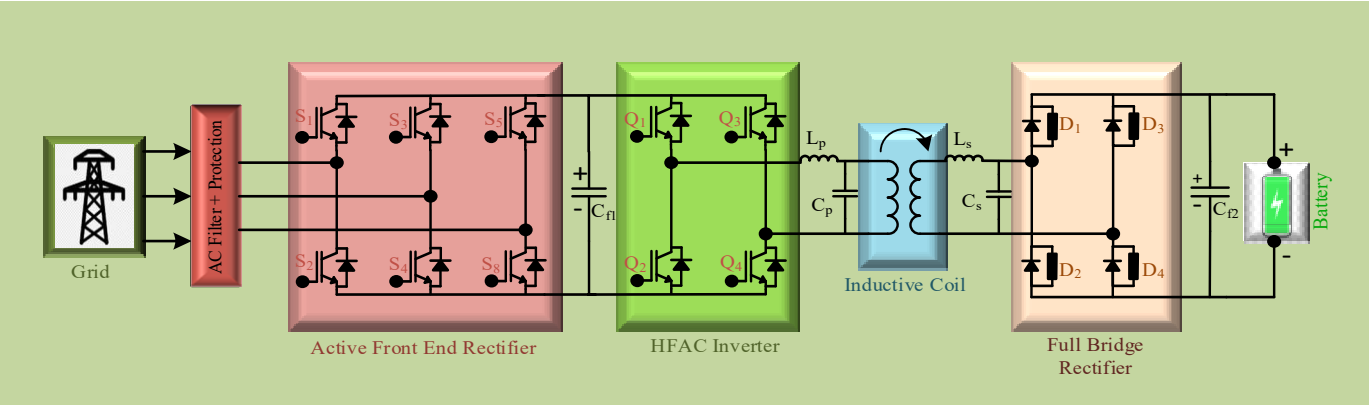


Fig.1. Charging topology for electric vehicles, showing primary and secondary charging stations

III. MODELLING

The vast majority of WPT designs employ planar coils. There are two distinct possibilities: circular and rectangular coils. The parameters of the circular spiral coils used in this study are depicted in the figure below. The LCR meter can be used to measure the self-inductance, resistance, and diameter of a circular spiral coil with and turns, respectively by connecting the terminals to the LCR meter as shown in Fig (2).

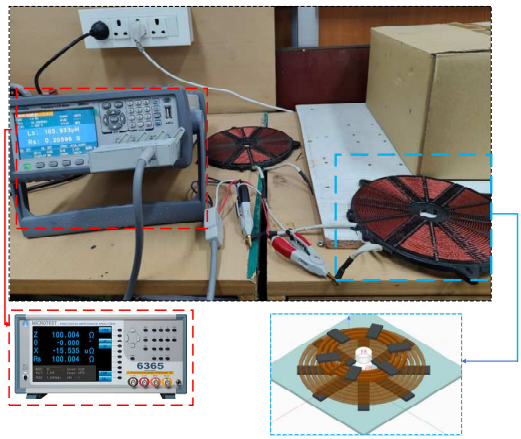


Fig.2 Measurement Setup by LCR Metre SM6023

TABLE.1.-COIL PARAMETERS

COIL PARAMETERS	VALUE
Roadside Winding Turns	$N_1 = 28$
Vehicle side Winding Turns	$N_2 = 28$
Diameter of Roadside Coil	$D_1 = 19cm$
Diameter of Vehicle side Coil	$D_2 = 19cm$
Self-Inductance of Roadside Coil	$L_1 = 155.243\mu H$
Self-Inductance of Vehicle side Coil	$L_2 = 110.313\mu H$
Resistance of Roadside Coil	$R_1 = 0.23213\Omega$
Resistance of Vehicle side Coil	$R_2 = 0.20755\Omega$
Operating frequency	$f_s = 30kHz$

Table (1) presents an overview of the coil parameters, including the number of turns in the roadside coil, its diameter, self-inductance, resistance, as well as the number of turns in the vehicle side coil, its diameter, self-inductance, resistance, and the operating frequency also shown in Fig (3).

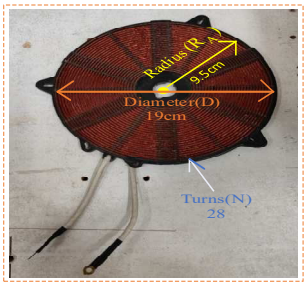


Fig.3 Front View and Specification of coil

Mono resonant compensation corrects system resonant with one frequency. Multi-resonant compensation systems with reactive components worked well. Our simulation tweaks basic LC topology. This work examines series-series compensation as shown in Fig (4) and LC topological parameters as shown in Fig (5). We study the development of these parameters by considering the roadside (V_1) and vehicle side (V_2) voltages and currents (I_1 and I_2).

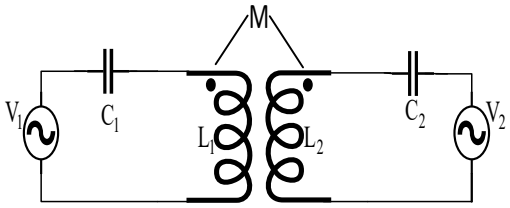


Fig.4 Series-Series mono-resonant Compensation

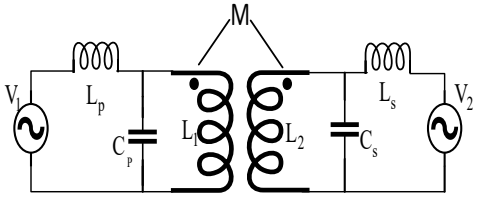


Fig.5 LC based multi-resonant Compensation

In order to determine parameters, a mesh-based analysis is first performed, after which the input impedance is inferred, and only components that are entirely resistive require their reactive component terms equal to zero.

$$V_1 = ((R_2 + j\omega L_1 + \frac{1}{j\omega C_1})I_1 - j\omega M I_2 \qquad (1)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:11:26 UTC from IEEE Xplore. Restrictions apply.

$$V_2 = ((R_2 + j\omega L_2)I_2 - j\omega M I_1 = (R_L + \frac{1}{j\omega C_2})I_2 \quad (2)$$

From Eq(1),(2), The Z_{in} and ratio between I_1 and I_2 as follows:

$$\frac{I_2}{I_1} = \frac{j\omega M}{j\omega L_2 + R_2 + R_L + \frac{1}{j\omega C_2}} \quad (3)$$

$$Z_{in} = R_1 + j(\omega L_1 - \frac{1}{\omega C_1}) + \frac{\omega^2 M^2}{j\omega L_2 + R_2 + R_L - \frac{j}{\omega C_2}} \quad (4)$$

As a result, the series-series adjusted parameters are able to be determined using Equation(5)

$$\omega_0 = \frac{1}{\sqrt{L_1 C_1}} = \frac{1}{\sqrt{L_2 C_2}} \quad (5)$$

Multi-resonant compensation parameters as the process:

$$L_p = 1mH \quad (6)$$

$$C_p = 0.18\mu F \quad (7)$$

$$L_s = 1mH \quad (8)$$

$$C_s = 0.255\mu F \quad (9)$$

Parameters on both the vehicle and road sides of compensation are determined and shown by Equation (6), (7),(8)and(9)

A. MUTUAL INDUCTANCE CALCULATION

Consider two circular spiral coils aligned in a way as shown in the Fig (6) with radius R_p and R_s and the distance between them from origin is h. As for analysis we are considering points A and B as its geometrical view is shown in Fig (7).

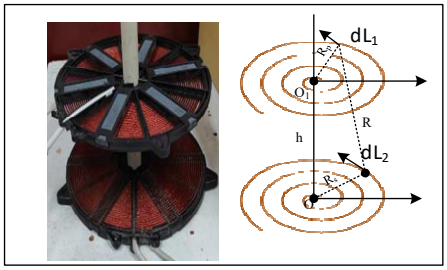


Fig.6 Planar spiral coil with perfect alignment

According to Neuman's formula, the mutual inductance of a planar spiral coil calculated using the magnetic vector potential method is as follows:

$$M = \frac{\mu_0}{4\pi} \oint_{L_1} \oint_{L_2} \frac{dL_1 \times dL_2}{R} \quad (10)$$

where μ_0 is the air permeability, dL_1 and dL_2 are line elements and R is the spacing between them.

Also, Line tangential elements at an arbitrary point A, B using the cylindrical coordinates system can be described as:

$$dL_1 = -R_p \sin\theta_1 d\theta_1 \hat{x} + R_p \cos\theta_1 d\theta_2 \hat{y} \quad (11)$$

$$dL_2 = -R_s \sin\theta_1 d\theta_1 \hat{x} + R_s \cos\theta_1 d\theta_2 \hat{y} \quad (12)$$

where R_p and R_s represents the radius of primary and Secondary coil with respect to origin

$$dL_1 \cdot dL_2 = R_p R_B \cos(\theta_2 - \theta_1) d\theta_1 d\theta_2 \quad (13)$$

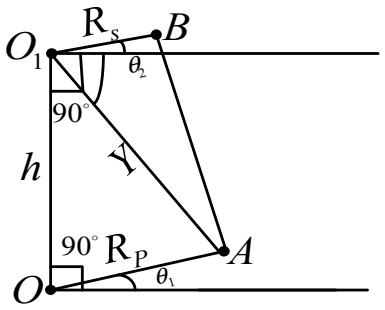


Fig.7 Geometry view for analysis

Using the cosine rule in triangle OO_1A, we can determine that Y is the distance to point AO_1

$$Y = h^2 + R_p^2 - 2hR_p \sin\theta_1 \quad (14)$$

By the triangle geometry rules

$$\angle AO_1B = \theta_2 - \theta_1 \quad (15)$$

Once more, we may use the cosine rule in triangle O_1BA to get the distance R between dL_1 and dL_2

$$R^2 = Y^2 + R_s^2 - 2YR_S \cos(\theta_2 - \theta_1) \quad (16)$$

$$R^2 = h^2 + R_p^2 - 2hR_p \sin\theta_1 + R_s^2 - 2R_S \cos(\theta_2 - \theta_1)\sqrt{h^2 + R_p^2 - 2hR_p \sin\theta_1} \quad (17)$$

As θ_1 and θ_2 is very small ,so $\sin(\theta_1)$ and $\sin(\theta_2)$ is very small and neglected but cosine quantities at small angle is comparable ,Hence the above equation after this approximation becomes as shown in Eq(18)

$$R^2 = h^2 + R_p^2 + R_s^2 - 2R_s R_p \cos(\theta_2 - \theta_1) \quad (18)$$

Substituting Equation (13) in Eq(10),the mutual inductance of perfectly aligned spiral coil is obtained in:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:11:26 UTC from IEEE Xplore. Restrictions apply.

$$M = \frac{\mu_0}{4\pi} \iint_{2\pi N} \frac{R_p R_B \cos(\theta_2 - \theta_1) d\theta_1 d\theta_2}{R} \quad (19)$$

Substituting Equation (17) in Eq (19), the mutual inductance of perfectly aligned spiral coil is obtained in:

$$M = \frac{\mu_0}{4\pi} \int \int_{2\pi N} \frac{R_p R_B \cos(\theta_2 - \theta_1) d\theta_1 d\theta_2}{\sqrt{h^2 + R_p^{\,2} - 2hR_p \sin\theta_1 + R_s^{\,2} - 2R_S \cos(\theta_2 - \theta_1)} \sqrt{h^2 + R_p^{\,2} - 2hR_p \sin\theta_1}} \quad (20)$$

Substituting our coil parameter in Eq (20) as shown in Table (1) value of Mutual Inductance is obtained as shown in Eq (21).

$$M_{accurate} = 39.33 \mu H \quad (21)$$

Substituting our coil parameters in Eq (19) as shown in Table (1), the value of Mutual Inductance is obtained as shown in Eq (22).as mentioned in [14]

$$M = \frac{\mu_0}{4\pi} \iint_{2\pi N} \frac{R_p R_B \cos(\theta_2 - \theta_1) d\theta_1 d\theta_2}{\sqrt{h^2 + R_p^{\,2} + R_s^{\,2} - 2R_s R_p \cos(\theta_2 - \theta_1)}} \quad (22)$$

Evaluating double Integral for the above parameters and coil turns of the roadside and vehicle side and its calculation leads to the Eq (23).

$$M_{approx} = 40.323 \mu H \quad (23)$$

The evaluation of the double integral for the aforementioned parameters and its subsequent calculation results in Equation (24).

$$M = 39.33 \mu H \quad (24)$$

IV. VERIFICATION

A. MUTUAL INDUCTANCE MEASUREMENT METHOD

This mutual inductance measurement is performed using a Scientific Precision LCR Meter SM6023, and the results are compared to the calculated values. To measure mutual inductance, the following method is employed. First, the inductance is measured when the primary and secondary coils are connected in forwarding mode, as depicted in Fig (8), and then when they are connected in backward mode, as depicted in Fig (9) Their one-fourth subtraction results in the calculation of mutual inductance.

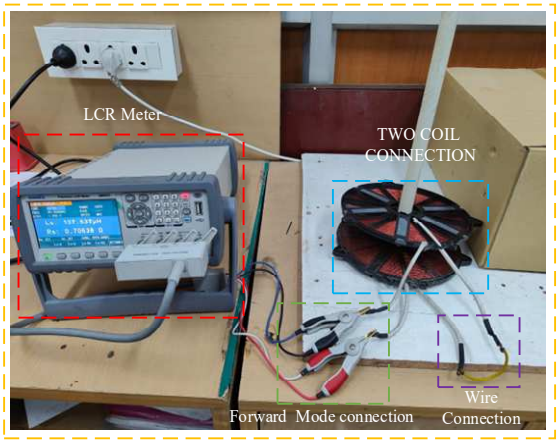


Fig.8 Forward mode Inductance test

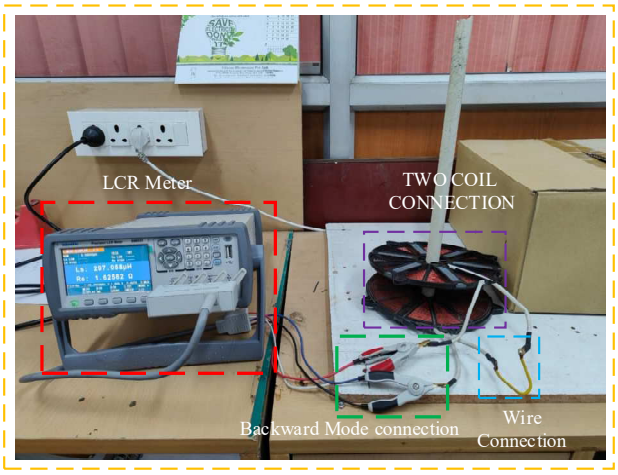


Fig.9 Backward Mode Inductance Test

The actual mutual inductance is derived by subtracting the forward connection inductance from backward connection inductance and dividing it by 4 leads to the equation as given in (20)

$$M = \left| \frac{L_f - L_b}{4} \right| \quad (25)$$

The forward mode Inductance Test, depicted in Figure 1, is accompanied by its corresponding readings as presented in Equation (26).

$$L_f - L_P + L_s + 2M = 137.637 \mu H \quad (26)$$

The Backward mode Inductance Test as shown in Fig.1.and its readings are shown in Equation (27).

$$L_b - L_P + L_s - 2M = 297.055 \mu H \quad (27)$$

Hence, utilizing this measurement technique, the value of mutual inductance as determined by Equation (29) closely approximates the value indicated in Equation (24).

$$M = \frac{297.055 - 137.637}{4} \quad (28)$$

$$M = 39.8545 \mu H \quad (29)$$

B. COUPLING FACTOR CALCULATION

The verification of the coupling coefficient, which is an essential aspect of evaluating mutual inductance, plays a crucial role in ensuring the efficiency and safety of wireless power transfer systems. The optimization of coil designs, alignment, and positioning for quality power transfer can be achieved through the identification of anomalies in the coupling coefficient.

$$K = \frac{M}{\sqrt{L_1 L_2}} \quad (30)$$

Substituting Equation (24) and L_1, L_2 through Table (1)

$$K = 0.2861 \quad (31)$$

The coupling factor, as determined by Equation (31), falls within an acceptable range, as demonstrated.in [5]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:11:26 UTC from IEEE Xplore. Restrictions apply.

V. RESULTS AND DISCUSSION

The Active Front End (AFE) system uses a 415(V)VPP three-phase AC power supply from a grid or generator. LCL filters minimise noise and harmonics. A 6-switch Bridge rectifier generates pulsing DC voltage from AC. IGBT bridges power this rectifier. The DC-link capacitor smooths 800V DC voltage. Voltage reference variations affect the AFE converter's voltage control loop. To maintain output voltage, this control loop minimises the difference between real and reference voltages. Figure 10 shows a reference voltage of 700V to 900V or 800V to 600V.

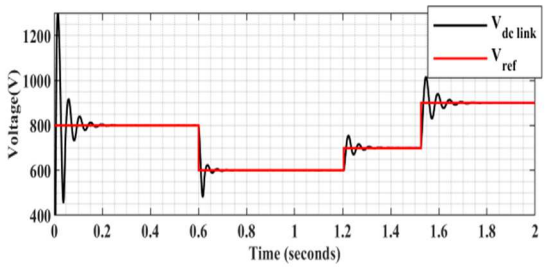


Fig.10 Output of AFE (V_{DC}) at different Voltage reference

A 30kHz, 50% duty cycle high-frequency alternating current (HFAC) inverter with an open-loop control system switch. The 800V input voltage (V_{DC}) inverter has reduced size and weight and faster response times. Due to the lack of input, an open-loop control system may have dynamic performance limits, although it can still be successful.

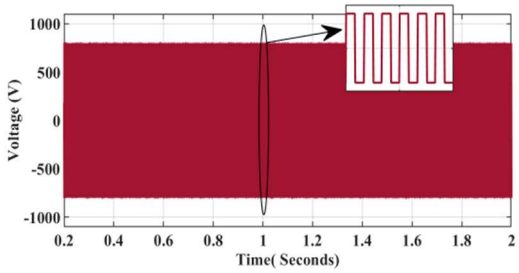


Fig.11 Output voltage of HFAC Inverter

Inductive Power Transfer (IPT), operates through the utilization of magnetic induction. An LC circuit is utilized to mitigate the effects of an (800V) V_{pp} HFAC square wave as Fig (11) by generating a sinusoidal AC voltage. The power is received by the primary coil and subsequently after compensation as Fig (13). transferred to the secondary coil, resulting in the generation of a high frequency voltage of (1000V) V_{pp} as Fig (14). Full Bridge Rectifier with Capacitive Filter power electrical circuit converts AC power to DC power. The rectifier's output is sent via a capacitive filter to charge a 48V lead-acid battery with a steady, high-quality DC output voltage.

The wireless EV charging topology shows a continuous sinusoidal pattern with a 50Hz frequency, indicating a stable and steady source of electricity. To improve Power Factor (PF) and reduce harmonic distortion, system monitoring is needed to reduce reactive power consumption in the current waveform. The AFE's well-regulated DC link voltage waveform reflects load or input grid supply changes as. as Fig (12) The battery voltage waveform shows rising and then stopping after it reaches 48Vas shown in Fig (12).

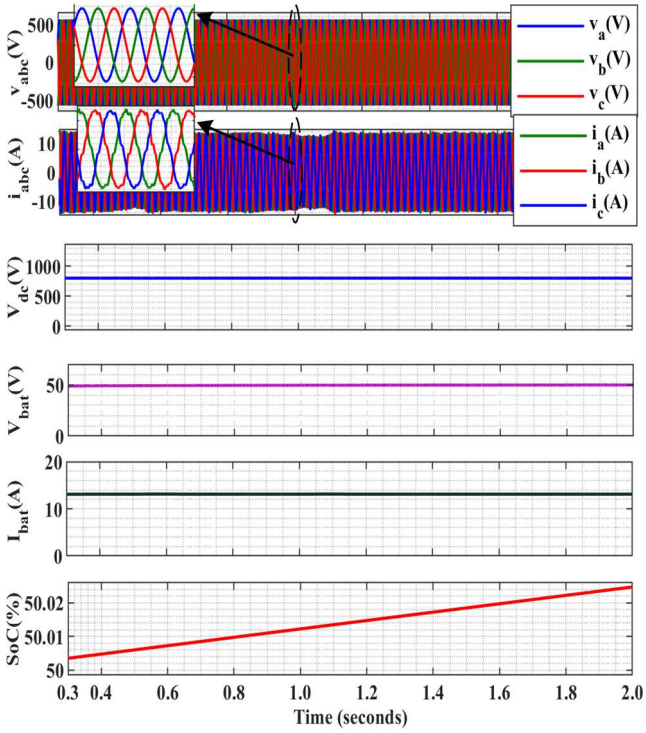


Fig.12 (a) Grid voltage(top), (b)grid current, (c) $V_{dc\ link}$ (AFE), (d) battery voltage, (e)battery current, (f) State of charge of battery(bottom)

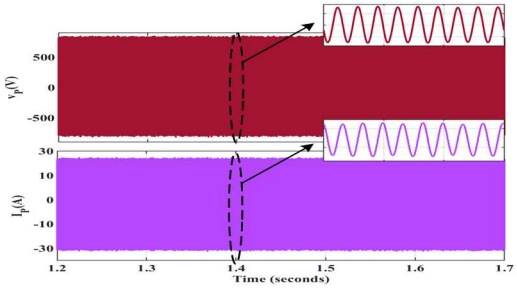


Fig.13 Voltage and current on the primary coil

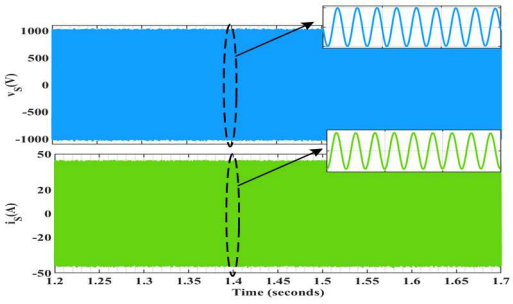


Fig.14 Voltage and current on the secondary coil

VI. CONCLUSION

In conclusion, the analysis of an LC compensated (WPT) charger, implementing thorough calculations of mutual inductance and assuming a perfect alignment with a mutual inductance value of M = 39.33, has exhibited considerable promise for charging batteries operating at a standard voltage of 48 volts, while meeting the requirements of conventional

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:11:26 UTC from IEEE Xplore. Restrictions apply.

battery specifications. By employing a grid-side voltage with a peak-to-peak value V_{pp} of 415 volts, the battery is effectively charged by the system, eliminating the need for a DC-DC converter on the load side. The charger is equipped with precise LC compensation, featuring primary and secondary inductance values of $L_p = 1mH$ and $L_s = 1mH$, respectively. Additionally, it incorporates capacitance values of $C_p = 0.18\mu F$ and $C_s = 0.25\mu F$, which effectively facilitate efficient power transfer while minimizing energy losses. Moreover, the utilization of an Active Front End (AFE) in the capacity of a grid-side AC-DC converter serves to accomplish genuine unity power factor correction (UPF), effectively aligning voltage and current waveforms.

APPENDIX

Control of AFE

Voltage control

$T_{C1} = 220\,ms$

$C_1 = 100\,mF$

$K_P = \dfrac{3 * C_1}{T_{C1}} = 1.5$

$K_I = \dfrac{3 * (20 * 10^{-3})}{T_{C1}} = 300$

I. *Current control*

$T_{C1} = 220ms$

$C_1 = 100mF$

$K_P = \dfrac{3 * C_1}{T_{C1}} = 1.5$

$K_I = \dfrac{3 * (20 * 10^{-3})}{T_{C1}} = 300$

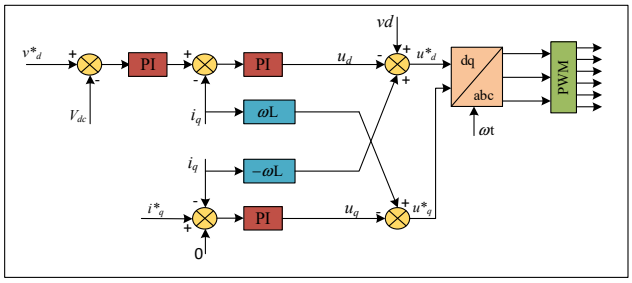


Fig.16. Voltage and Current loop Controlled Scheme

REFERENCES

[1] A. A. S. Mohamed, D. Allen, T. Youssef and O. Mohammed, "Optimal design of high frequency H-bridge inverter for wireless power transfer systems in EV applications," *2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC)*, Florence, Italy, 2016, pp. 1-6, doi: 10.1109/EEEIC.2016.7555646.

[2] S. Li, F. Li, R. Zhang, C. Tao and L. Wang, "Accurate Modeling, Design and Load Estimation of LCC-S based WPT system with a wide range of load," in *IEEE Transactions on Power Electronics*, doi: 10.1109/TPEL.2023.3279659.

[3] F. P. Wijaya, T. Shimotsu, T. Saito and K. Kondo, "A Simple Active Power Control for a High-Power Wireless Power Transmission System Considering Coil Misalignment and Its Design Method," in *IEEE Transactions on Power Electronics*, vol. 33, no. 11, pp. 9989-10002, Nov. 2018, doi: 10.1109/TPEL.2018.2799679.

[4] K. Fotopoulou and B. W. Flynn, "Wireless Power Transfer in Loosely Coupled Links: Coil Misalignment Model," in *IEEE Transactions on Magnetics*, vol. 47, no. 2, pp. 416-430, Feb. 2011, doi: 10.1109/TMAG.2010.2093534.

[5] A. Ramezani and M. Narimani, "Optimized Electric Vehicle Wireless Chargers with Reduced Output Voltage Sensitivity to Misalignment," in *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 8, no. 4, pp. 3569-3581, Dec. 2020, doi: 10.1109/JESTPE.2019.2958932.

[6] .N. Fu, J. Deng, Z. Wang, W. Wang and S. Wang, "A Hybrid Mode Control Strategy for LCC–LCC- Compensated WPT System With Wide ZVS Operation," in *IEEE Transactions on Power Electronics*, vol. 37, no. 2, pp. 2449-2460, Feb. 2022, doi: 10.1109/TPEL.2021.3108637

[7] Y. Li, J. Zhao, Q. Yang, L. Liu, J. Ma and X. Zhang, "A Novel Coil with High Misalignment Tolerance for Wireless Power Transfer," in *IEEE Transactions on Magnetics*, vol. 55, no. 6, pp. 1-4, June 2019, Art no. 2800904, doi: 10.1109/TMAG.2019.2904086

[8] L. Li, H. Liu, H. Zhang and W. Xue, "Efficient Wireless Power Transfer System Integrating with Metasurface for Biological Applications," in *IEEE Transactions on Industrial Electronics*, vol. 65, no. 4, pp. 3230-3239, April 2018, doi: 10.1109/TIE.2017.2756580.

[9] Y. Zhang et al., "Integration of Onboard Charger and Wireless Charging System For Electric Vehicles With Shared Coupler, Compensation, and Rectifier," in *IEEE Transactions on Industrial Electronics*, vol. 70, no. 7, pp. 7511-7514, July 2023, doi: 10.1109/TIE.2022.3204857.

[10] . A. A. S. Mohamed, C. R. Lashway and O. Mohammed, "Modeling and Feasibility Analysis of Quasi-Dynamic WPT System for EV Applications," in *IEEE Transactions on Transportation Electrification*, vol. 3, no. 2, pp. 343-353, June 2017, doi: 10.1109/TTE.2017.2682111.

[11] Y. Jiang, L. Wang, Y. Wang, J. Liu, M. Wu and G. Ning, "Analysis, Design, and Implementation of WPT System for EV's Battery Charging Based on Optimal Operation Frequency Range," *in IEEE Transactions on Power Electronics*, vol. 34, no. 7, pp. 6890-6905, July 2019, doi: 10.1109/TPEL.2018.2873222.

[12] S. R. Meher, Y. Choudhary and R. K. Singh, "An Optimal Wireless Battery Charger for Electric Vehicle using EF2 Inverter at 6.78 MHz," 2022 *IEEE Energy Conversion Congress and Exposition (ECCE),* Detroit, MI, USA, 2022, pp. 1-5, doi: 10.1109/ECCE50734.2022.9947370.

[13] A. Mishra and M. Singh, "A 10-kW Active Front End Rectifier Fed Dual Active Bridge Converter For EV Charging," 2023 *4th International Conference for Emerging Technology (INCET),* Belgaum, India, 2023, pp. 1-6, doi: 10.1109/INCET57972.2023.10170584.

[14] Hussain, I.; Woo, D.-K. Simplified Mutual Inductance Calculation of Planar Spiral Coil for Wireless Power Applications. *Sensors* 2022, *22*, 1537. https://doi.org/10.3390/s22041537.

[15] M. Ansari, G. Yadav and M. Singh, "A 3.3kW Modified LLC Resonant Converter for Grid-Tied EV System Under Wide Voltage Range," 2023 3rd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2023, pp. 1-6, doi: 10.1109/CONIT59222.2023.10205690.

[16] M. Ansari, G. Yadav and M. Singh, "A 3.3kW Modified LLC Resonant Converter for Grid-Tied EV System Under Wide Voltage Range," 2023 3rd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2023, pp. 1-6, doi: 10.1109/CONIT59222.2023.10205690.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on February 21,2024 at 06:11:26 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2022.3163277, IEEE/ACM Transactions on Computational Biology and Bioinformatics

VGGIN-Net: Deep Transfer Network for Imbalanced Breast Cancer Dataset

Manisha Saini, and Seba Susan

Delhi Technological University, Delhi, India

Abstract— In this paper, we have presented a novel deep neural network architecture involving transfer learning approach, formed by freezing and concatenating all the layers till block4 pool layer of VGG16 pre-trained model (at the lower level) with the layers of a randomly initialized naïve Inception block module (at the higher level). Further, we have added the batch normalization, flatten, dropout and dense layers in the proposed architecture. Our transfer network, called VGGIN-Net, facilitates the transfer of domain knowledge from the larger ImageNet object dataset to the smaller imbalanced breast cancer dataset. To improve the performance of the proposed model, regularization was used in the form of dropout and data augmentation. A detailed block-wise fine tuning has been conducted on the proposed deep transfer network for images of different magnification factors. The results of extensive experiments indicate a significant improvement of classification performance after the application of fine-tuning. The proposed deep learning architecture with transfer learning and fine-tuning yields the highest accuracies in comparison to other state-of-the-art approaches for the classification of BreakHis breast cancer dataset. The articulated architecture is designed in a way that it can be effectively transfer learned on other breast cancer datasets.

Index Terms—VGG16, pre-trained model, Inception module, transfer learning, deep learning, convolutional neural networks, fine tuning.

• *Manisha Saini is a research scholar in the Department of Computer science and Engineering, Delhi Technological University, Delhi, India, E-mail: manisha.saini44@gmail.com*
• *Seba Susan is a Professor in the Department of Information technology, Delhi Technological University, Delhi, India, E-mail: seba_406@yahoo.in*

1 Introduction

Cancer is a grievous health problem that affects both developed and developing countries [1]. Breast cancer is one of the most commonly found cancer in women. There are certain challenges faced while diagnosing cancer manually from the Histopathological biomedical images due to a high probability of inaccurate detection that might occur due to human error. In order to avoid the manual time consuming task for diagnosis of cancer, a computer-aided diagnosis system is required which automates the process efficiently. Various automated processes are already available to distinguish the Malignant and Benign cancer images [2]. Deep learning has further improved the performance of the automated process for the diagnosis of cancer at the early stage. The role of deep learning in various applications has increased tremendously with the recent advancements in computational hardware accelerators and parallel processing. Health care and biomedical are popular fields where deep learning has shown remarkable improvements [3].

The core of the deep learning architecture is the convolutional neural network (CNN) [4]. Basic CNN architecture consists of various convolutional building blocks stacked together for automating the process of feature extraction and classification, unlike machine learning where both the steps are performed separately. Broadly CNN networks are categorized into (i) networks which are trained from scratch, and (ii) the pre-trained networks [5] which are trained beforehand on a large dataset facilitating the transfer of knowledge from one domain into another.

A common problem associated with real-world datasets is the class imbalance problem. It is a challenging task to tackle the imbalanced dataset due to the disproportionate distribution of samples of different classes. The BreakHis breast cancer dataset is an instance of an imbalanced dataset with the Malignant samples outnumbering the Benign samples. The imbalanced nature of this dataset [6] brings in several challenges as the class imbalance problem causes several incorrect predictions. So there is a need to have a model which can handle the class imbalance situation effectively and will be able to correctly detect cancerous patterns from bio-medical data. In order to tackle this problem, we have proposed a novel deep learning based architecture incorporating transfer learning in this paper. The transfer learning approach using pre-trained networks has the advantage of transferring the learned weights from an architecture which is trained on another domain to the biomedical domain, which will ultimately reduce the training time and computational cost required to train the model from scratch. The contribution of the paper can be summarized as: (a) Successful design of novel deep network architecture using transfer learning approach to solve

https://github.com/SainiManisha/VGGIN-Net

class imbalance problem in breast cancer datasets. The proposed architecture is created by combining the relevant layers from VGG16 pre-trained network (layers till block4 pool layer) along with the naïve Inception module in combination with flatten, batch normalization and dense layers. Also, certain regularization techniques have been infused in the proposed model such as data augmentation and dropout which overall helps to reduce the overfitting to a great extent. (b) The proposed network has been successfully tried and tested on images of different magnification factors: 40X, 100X, 200X and 400X, establishing the network's invariance to size and scale of the image. The results indicate an overall improvement in the classification performance in comparison to other state-of-the-art approaches. (c) We have analyzed the effect and significance of block wise fine-tuning on the proposed deep transfer neural network architecture which has substantial impact on the classification performance. (d) The proposed 24 layer network architecture has been articulated by integrating the right combination of layers in well-ordered way to reduce the computational complexity involved. The formulated network architecture is designed in the way that it can be successfully transfer learned on other breast cancer datasets.

The various sections of the paper are summarized as follows. Section 2 describes the state of the art in the fields of machine learning and deep learning for breast cancer diagnosis. In section 3, the architecture and learning methodology used in the proposed work are discussed in detail. Further, in Section 4, a discussion regarding the datasets used in the experimental task along with the implementation details have been presented. In section 5, the final analysis and discussion on results related to the classification task performed and comparison with other state-of-the-art networks is presented. The final conclusion is given in the last section of the research paper.

2 Related work

Various machine learning and deep learning based networks have been proposed for detecting cancerous patterns from cell images [7]. In [46], authors had proposed an appropriate feature-classifier combination for multi- class imbalanced datasets by extracting the deep features using visual codebook generation along with the non-linear Chi2 SVM classifier. A number of studies have been reported in the literature using CNN for classification of medical images. Spanhol et al. (2016) used patches, extracted by applying the sliding window mechanism in order to efficiently train CNNs [8]. Feng et al (2018) had also focused on the patch based learning from images so in order to emphasize on the reduction of complexity of the network by reducing the number of parameters [9]. Bayramoglu et al. (2016) proposed single-task and multi- task CNN models for the classification of BreakHis Histopathological dataset [10]. Bardou et al. (2018) had compared various configuration combinations of CNN with other traditional approaches. They have also emphasized the role of data augmentation operations on the performance of the model. From the results analysis, it was validated that the deep features extracted using CNN outperforms the handcrafted features [11].

In order to tackle the class-imbalance problem which occurs due to an unequal distribution of samples in the dataset, several methods have been suggested in data mining such as undersampling, oversampling and hybrid sampling strategies [12],[13]. Data augmentation operations can be applied on minority class and majority class, or it can be applied separately on the minority class [14] [43] in order to make the samples of the minority class equivalent to the majority class. In [43], the authors had increased the number of samples in the minority class using the Deep Convolutional Generative Adversarial Network (DCGAN) by synthetically generating the fake samples, and the augmented dataset was learned using a modified VGG16 based architecture. In previous work, comparison of BOVW, deep neural networks and data augmentation for the imbalanced computer vision dataset was presented in [15]. It was observed from the study that the deep neural network mitigates the effect of class imbalance in the most effective manner. Traditionally, the methods that are used to tackle the class-imbalance problem modify the data distribution, which can be challenging for medical datasets as it may cause loss of relevant information. Ding et al. (2017) demonstrates that the use of very deep networks (models with more than 10 layers) improves the network training process and achieves better rate of convergence for imbalanced datasets [45]. The authors experimentally validate the claim by using deep network architectures which are trained for 100 epochs in comparison to shallower networks. There are other relevant approaches also proposed by researchers to deal with the imbalance situation such as Abbas et al [44] had proposed a Decompose, transfer and compose (DeTraC) model using the concept of class decomposition within CNN approach, which helps to learn the class boundaries effectively. The classification task is performed after ensuing the error correction criteria applied to the softmax layers of the network which indeed had improved the classification performance.

Different pre-trained networks have been used before by many researchers to detect cancer patterns. Rakhlin et al. (2018) had worked upon the ICIAR 2018 Grand Challenge on Breast Cancer Histology Images by extracting deep features from the pre-trained networks, that were trained using the gradient boosted trees classifier [16]. Deniz et al. (2018) concatenated the features extracted from various layers of AlexNet and VGG16 pre-trained models. The deep features extracted from these layers were used in conjunction with support vector machine classifier for the classification of Benign and Malignant cancer images [17]. Gupta and Bhavsar (2018) had proposed a sequential framework which consists of deep multi-layer features extracted from fine-tuned DenseNet on BreakHis dataset. The deep features extracted from multiple layers were given to XGBoost classifier [18]. In our proposed transfer learning approach, we have created a new deep architecture by concatenating pre-trained layers at lower level with trainable layers at

https://github.com/SainiManisha/VGGIN-Net

higher level. The higher layer features are learned specific to the breast cancer dataset. This makes our approach distinctive from previous works since we focus on transferring general features from the lower layers of pre-trained large convolutional models while the higher-level features are learned specifically from the cancer dataset.

3 PROPOSED DEEP TRANSFER NETWORK: VGGIN-NET

A novel deep neural network architecture is proposed by concatenating the VGG16 architecture layers [19] with a single naïve Inception block [20], as shown in Figure 1. Inception block is a module which is popularly used in Inception architecture (a.k.a. GoogLeNet) [20]. We have presented a transfer learning approach involving the novel deep neural network architecture, that we have named as VGGIN-Net, which is formed by freezing and concatenating all the layers till block4 pool layer of the VGG16 pre-trained model (at the lower level) with the layers of a randomly initialized (using Glorot uniform distribution) naïve Inception module (at the higher level). Further, several randomly initialized higher layers are added such as the dense layer along with batch normalization, flatten, and dropout layers. The proposed architecture is created by concatenating the layers as shown in Figure 2. The new deep network is now trained on the breast cancer dataset.

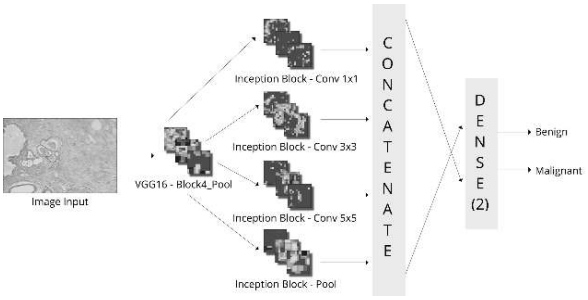


Figure 1. Proposed deep network architecture VGGIN-Net showing the lower layers till block4 pool layer of VGG16 pre-trained network and the higher layers comprising of naïve Inception module and the dense layers.

The 24-layer architecture is constructed as shown in Figure 2 by first stacking the VGG16 layers, starting from the VGG16 pre-process layer till the block4 pool layers. The 224 x 224 x 3 image is given to the VGG16 pre-trained network as input. The reason for considering the features till block4 pool layers from the VGG16 pre-trained model is to extract the most relevant bottleneck features. Also, the consideration of features beyond the block4 pool layer would only increase the computational difficulties with improvement in the performance of the model, as validated by our experimental results. Further, the obtained relevant bottleneck features till block4 pool layer of VGG16 pre-trained network are concatenated with the layers of naïve Inception block. The naïve Inception block consists of convolutional layers having filters of sizes of 1x1, 5x5 and 3x3, with each layer having stride of size 3x3 and ReLU (Rectified linear unit) activation function [21], with addition of max pooling 2D layer.

In this paper, we have considered the goodness of both the pre-trained models (VGG16 and Inception) to create a more robust architecture which effectively resolves the class imbalance problem. The use of VGG16 pre-trained layers in the initial stage of our network was motivated by the fact that VGG16 architecture achieves good accuracy for most of the image classification problems and this network is efficient in dealing with images of different scales and magnification factors. This is particularly useful for our study as one of the target datasets that we experiment with, is the BreakHis dataset which consists of images of varying magnification factors. Throughout our study, we do not introduce skip connections or depth separable convolutions as they are well beyond our scope of this study. The idea is to improvise the goodness of popularly used existing architectures through a sequentially stacked layer model and use of multi-level features that adapt better to histopathological datasets.

There are certain crucial deployment challenges faced in the VGG based architecture which motivated us to modify the VGG16 architecture. As inspired from the previous works [22],[23],[24],[25] we have modified the VGG16 architecture so as to overcome the deployment challenges that comes with VGG-Nets. Such computational challenges are prevalent even on powerful single-GPU systems (Graphical Processing Units) due to its large memory footprint. The sequential ConvNet, VGG16 bears a large number of parameters (140 million) due the presence of multiple convolutional layers of varying receptive fields, hence, it can become inefficient for inference at test time. Due to the large number of parameters, VGG16 network is also prone to vanishing gradient problem. The presence of three fully connected layers in the original VGG16 architecture is primarily responsible for the major bulkiness of the model. So we have extracted the relevant deep features uptill block4 of the pre-trained model and removed all the layers after that which added extra complexity and computation in the proposed architecture. Further, to address the shortcomings seen in the VGG16 architecture, we have added the naïve Inception block as an additional block in the proposed architecture. The GoogLeNet incarnation of the Inception architecture uses multiple auxiliary classifiers connected to intermediate layers to tackle the vanishing gradient problem. In our case, any auxiliary loss has not been used to train the inception block because of the presence of lesser number of images in the currently used dataset in comparison to the large-scale ImageNet. The reason for the addition of a single Inception block in the higher layers of the proposed architecture is that it would not require auxiliary classifiers and the model can converge by itself using a single loss only. Also, it would be less computationally expensive to add a single Inception module instead of addition of multiple similar modules. The main idea behind adding the naïve Inception module in the higher end of our network is to cover a larger region of a convoluted image yet preserving the finer details. The naïve inception block is specifically engineered to convolve in parallel such that accurate detailing is

https://github.com/SainiManisha/VGGIN-Net

possible through 1x1, 3x3, 5x5 convolutions (64, 128, 32 filters were used respectively). The goal of adding the naïve module is to increase the CNN's learning ability and abstraction of complex filters which was also found to be a drawback in VGG-based architectures. Moreover, the advantage of the Inception architecture is that it is able to perform well even with a single fully connected layer [20].

Our consideration of the naïve Inception block as a suitable choice along with a single dense layer makes the architecture less computationally expensive. A single dense (fully connected) layer with softmax activation function is present at the output to learn the proposed network architecture to deal with the higher end linear features and to find the probability of occurrence of the image belonging to each class for the two class classification problem (i.e. Benign and Malignant). Further, batch normalization, flatten, and dropout layers are added to enhance the network performance.We refrain from using multiple batch normalization layers (one batch normalization per convolutional layer) as per inspiration from [26] as a single batch normalization should suffice when layers are being concatenated as in case of the Inception module. Flatten layer is added after batch normalization to reduce the features into one dimension of size 144256. Dropout regularization [27] with a rate of 0.4 is added after the flatten layer which in turn helps to avoid overfitting problems.

Our transfer network thus, facilitates the transfer of domain knowledge from the larger ImageNet object dataset [28] to the smaller breast cancer dataset in the lower layers itself, and learns the higher layer features specific to cancer images. In doing so, we follow the guidelines of Yosinski et al. (2014) [29] who propounded the theory that when the base and target datasets are dissimilar, the use of pre-trained weights alone may degrade the performance. It is essential that the higher level features should be specific to the target dataset instead of the base dataset. Another regularization technique involved in the proposed architecture,

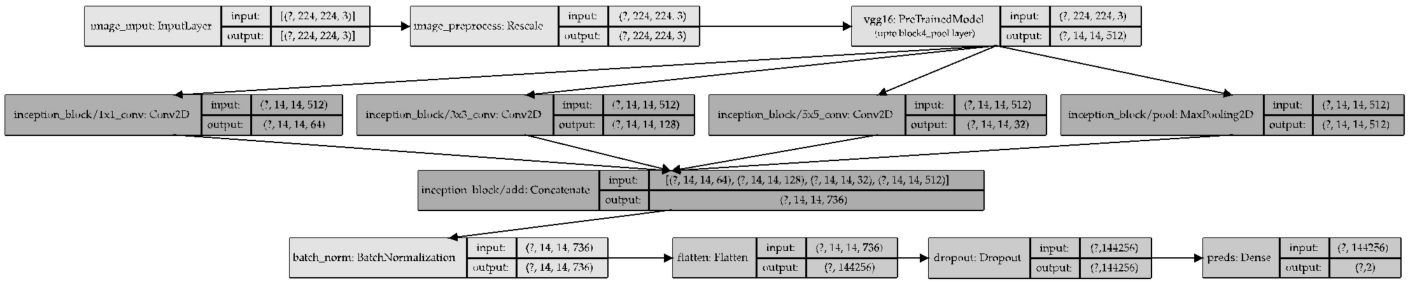


Figure 2. Proposed model architecture VGGIN-Net formed by concatenating VGG16 layers uptill block4 layer with the naive Inception module along with dense layer used for the classification

specifically at data-level, is data augmentation which is applied on the training dataset in order to synthetically increase the number of samples and also to improve the overall performance of the network by reducing the fitting problem [30]. The typical CNN training process employing data augmentation would include on the fly generation of random image samples across training mini-batches by use of affine transformations. In the proposed network, we have applied certain data augmentation operations as inspired from [31] on both the classes (Benign and Malignant). Figure 3 displays some randomly generated samples after applying data augmentation operations. The data augmentation operations applied on images include: (a) random rotation within range of 20 degrees, (b) random width and height shift operation with range 0.2 i.e. translation of the images both horizontally as well as vertically by number of pixels less than or equal to 20% of the actual image dimensions, (c) random horizontal and vertical flip, combined with (d) random shear and random zoom operations with the same range as that of translation. These values were determined after lots of experimental trials in order to maximize performance. To improve the regularization of our network further we make use of random crops [31] which helps the network to learn even better due to the translation invariance property of convolutional networks. The image patches are resized to 224 x 340 using bilinear resizing and then randomly cropped into patches of 224 x 224. At inference (test) time, central crop was used.

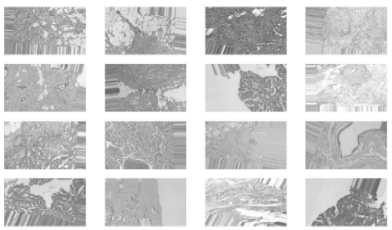


Figure 3. Illustration of various data augmentation operations applied on applied on the BreakHis dataset

Fine tuning approach has further been adapted on the proposed deep transfer learning architecture, inspired from several works in literature. In the case of the DeTrac approach, Abbas et al [43] had focused on the relevance of applying fine-tuning on different architectural blocks of pretrained CNN. Similarly, Sharma and Mehra (2018) had emphasized the role of transfer learning, full training, and fine-tuning of several pre-trained networks for the medical image dataset [32]. From the analysis it was found that VGG16 pre-trained network features with logistic regression as classifier gives best performance amongst other combinations of VGG19 and ResNet-50 pre-trained network with regression. The same authors

https://github.com/SainiManisha/VGGIN-Net

had further extended their work in [33] and elaborated on the role of layer wise fine-tuning and had presented the in-depth study of layer wise fine-tuning on AlexNet for the BreakHis dataset. From the study they had depicted that different magnification factors had influence in selecting the appropriate layers to be fine-tuned in the network architecture. Kandel and Castelle (2020) [34] had also conducted an extensive comparative analysis of block wise and complete fine-tuning of various pre-trained networks. The pre-trained networks used were: VGG16, VGG19 and Inception on histopathological image dataset. From the analysis it was found that fine-tuning of complete pre-trained network might not be the ideal choice in all the situations while considering different magnification factor images. All these previous works motivated us to apply block-wise fine tuning approach on the proposed network.

Our research work proposes a novel approach by combining modified VGG16 architecture with naïve Inception block to tackle imbalanced problem in breast cancer classification. The same has been empirically validated by conducting extensive experimentation along with ablation study to prove the veracity of our claim that our proposed architectural combination is able to solve the breast cancer classification task effectively by proposing the explainable and less computationally costly architecture. By adjusting the sequence and right combination of appropriate layers in the proposed architecture we are able to obtain the competent architecture to enhance the performance of the classifier that can be utilized for transfer learning on any other breast cancer dataset. The modified VGG16 architecture is chosen in such a way to resolve the deployment issues associated with original VGG-Nets by reducing the number of dense layer to one and extracting the appropriate features from suitable layers. In addition, we have introduced the naive Inception block with the batch normalization layer to address the vanishing gradient problem, and data augmentation, regularization and fine-tuning techniques have been used for improving the prediction performance.

4. EXPERIMENTAL SETUP

4.1 Datasets

Two different datasets are considered in the experimental task: BreakHis and Breast-Histopathological-Images dataset. BreakHis dataset is used to train the proposed network. The trained network also supports transfer learning as validated by the classification task performed on the Breast-Histopathological-Images dataset.

4.1.1 BreakHis Histopathological dataset

BreakHis dataset [35] used for current study is a publicly available dataset consisting of 7909 high resolution breast cancer images belonging to two different classes. i.e. Benign and Malignant, and composed of different magnification factors, 40X, 100*X*, 200*X* and 400*X* as illustrated in Table 1. The dataset also contains labels comprising of eight sub types belonging to four benign tumor types: adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenona (TA); and four malignant tumor types: carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC) and papillary carcinoma (PC). Skewed classes are observed in scenarios where the number of observations belonging to one class is significantly lower than other class. This is usually referred to as the class imbalance problem; this problem is prevalent in the BreakHis dataset as evident from Table 1 where the number of Malignant samples is far more than the number of Benign samples. For conducting the experimental task, 70:30 split is selected. 60% samples are selected for the training, 10% samples are kept for validation and remaining 30% samples are for testing purpose. The original size of images present in the BreakHis dataset is 700 x 460.

TABLE 1
SAMPLE DISTRIBUTION OF BENIGN AND MALIGNANT IMAGES IN THE BREAKHIS DATASET WITH RESPECT TO THE DIFFERENT MAGNIFICATION FACTORS

Class	Magnification Factor			
	40*X*	100 *X*	200 *X*	400 *X*
Benign	625	644	623	588
Malignant	1,370	1,437	1,390	1,232
Total No of Images (7909)	1,995	2,081	2,013	1,820

4.1.2 Breast-Histopathological-Images dataset

We have also validated our deep transfer network on another breast cancer dataset. For another set of experiments we have considered the Breast-Histopathological-Images dataset [36]. The Breast-Histopathological-Images dataset comprises of 277524 image samples having microscopic views of breast cell specimens at 40X magnification factor. Each image patch extracted from whole slides is of size 50 x 50 and all experiments are evaluated using a similar stratified split such that 30% of the total samples are kept for testing purpose and remaining 60% for training and 10% for validation. The dataset tries to address growing challenges in detecting invasive ductal carcinoma (IDC), which is the most common type of breast cancer. A highly imbalanced class distribution is observed in this dataset with 198738 IDC –ve images and 78786 IDC +ve images. It is interesting to note in this particular dataset, images from the IDC -ve class are in majority in comparison to the number of images present in the IDC +ve classes.

4.2 Experimental setup

All the experiments were conducted on Google Cloud Platform using a single Compute Engine VM instance with dual-core Intel Xeon CPU (2.00 GHz) and 8GBRAM, and a NVIDIA Tesla T4 GPU accelerator with 16 GB memory. For performing all our experiments, we used the TensorFlow v2.3.0 framework with the help of Keras API [37] using Python v3.8.9. The source code containing the code for our implementations is available in our GitHub

https://github.com/SainiManisha/VGGIN-Net

repository.[1] The different hyperparameters were selected so as to maximize the performance. (i) Adam optimizer, used with learning rate initially assigned as 0.001; the Adaptive Momentum optimization automatically adjusts the learning rate for further training. (ii) The loss function used was categorical cross entropy. (iii) The training batch size was set to 128 with a net budget of 100 epochs. The proposed architecture is fine-tuned for four different magnification factors. While fine tuning the network, the learning rate is significantly reduced so as to make sure that any large gradient updates would not cause the network to abruptly change any of the pre-trained weights. The training process is conducted with the help of a simple learning rate schedule where we exponentially decay the learning rate after 15 starting warmup epochs. During the fine tuning process the network is trained for a total of 50 sweeps (epochs). The warmup steps linearly increase the learning rate from 1e-5 to 5e-5 which is further exponentially decayed by a factor of 0.8.

In order to verify our claim that the proposed network architecture can further support transfer learning based tasks on other target histopathological images dataset, we use the weights of our VGGIN-Net trained on BreakHis 40X dataset with some amount of fine tuning to classify IDC +ve and –ve images from Breast-Histopathology-Images dataset. The weights chosen were from the 40X magnification factor as the target dataset also consists of images scanned at 40X zoom factor. The hyperparameters for training the terminal dense layer for the new classification is similar to our other experiments except that we use the Adam optimizer with a learning rate of 0.001 as any learning rate higher than that would affect the performance of the classifier.

5 Results and Discussion

5.1 Performance Analysis

In our study, we have done the comparative analysis of the proposed architecture with few state-of-the-art deep learning approaches as well as popularly used CNN architectures. For this curated set of architectures we primarily apply a transfer learning approach based on weights pre-trained on the ImageNet dataset. This is due to the fact that our target imbalanced classification dataset contains much less samples in comparison to large scale datasets (a few million images) which is almost always required to effectively train a large ConvNet from scratch. Initially, we experiment with VGG architecture using the well-known VGG16 network proposed by Zisserman et al. [19] (2015). We also used the GoogLeNet incarnation of the Inception architecture (as per work done in [20] by Szegedy et al. (2015)) as well as deep residual network ResNet-50 [26] proposed by He et al (2018) to classify histopathological images from the imbalanced BreakHis dataset. These methods when applied using transfer learning serve as effective baseline methods. So we have evaluated the performance of popular deep learning approaches i.e., VGG16, GoogLeNet and ResNet-50 in Table 2. We have also done the comparative analysis between these methods and our proposed deep transfer network, named VGGIN-Net, obtained after considering the features extracted till block4 pool layer of VGG16 with a single dense layer, and also with the addition of an Inception block and a single dense layer which is the proposed architecture. The VGG16 architecture had been modified with a single dense layer but after the addition of a naïve Inception block, the same network architecture had shown tremendous improvement in the results. From the results tabulated in Tables 2 to 9, it is observed that VGGIN-Net shows remarkable improvement in terms of accuracy, F1 score, IBA and GMean. ROC curve analysis with its AUC is also shown in Figure 5 to validate the proposed approach.

In Table 3, comparative analysis with state-of-the-art methods is shown for the BreakHis dataset using accuracy as the evaluation parameter with scores reported across several runs. Hence, it is evident that our proposed network with and even without fine tuning shows remarkable improvement in results.

To emphasize the veracity of our claim that the proposed network architecture helps to tackle the class imbalance problem, certain experiments were conducted. A comparative analysis illustrates the use of various well known approaches that deems to solve the class imbalance problem i.e., with undersampling and oversampling techniques. We observe from the comparisons to sampling experiments in Table 4 that the proposed architecture itself is able to tackle the class imbalance problem by itself without requirement of any sampling technique.

Extensive experiments were conducted related to the block-wise fine-tuning technique applied on the proposed network. It is observed from the analysis that the block-wise fine tuning operations have shown significant improvement in the performance as depicted in Table 5. It is evident that different fine tuning combinations are found suitable for different magnification factors. For 40X, fine tuning of block3, block4 and Inception block seems to be an ideal choice, whereas in case of 400X, fine tuning of block4 and Inception block was only found to be the perfect fit. Fine tuning of the complete network was found to be ideal in case of 100X and 200X magnification factor images. It can be inferred from the results that complete fine-tuning of the network is not always the perfect choice for different magnification factor images as different block wise fine tuning combinations also be deemed to be suitable in certain scenarios. Figure 4 depicts the validation accuracy and loss corresponding to the four magnification factors. It is clearly visible that with the help of suitable block-wise fine tuning, the network training improves its anytime performance, and learning curves get more and more stable. The proposed approach along with fine tuning has shown significant improvements in the classification performance. For $40X$, $100X$, $200X$ and $400X$ magnification factors of the BreakHis dataset, the best obtained accuracies are 98.51%, 97.53%, 96.688% and 95.528% respectively. So in a nutshell, it is notable to mention that our work

[1] https://github.com/SainiManisha/VGGIN-Net

https://github.com/SainiManisha/VGGIN-Net

demonstrates that single branch models can converge quite well and our work is not intended to deal with training of increasingly complex residual models. Rather, we are aimed at building a simple model with reasonable depth and favorable accuracy that can be simply implemented using basic architecture blocks (like convolution, ReLU, max pooling, etc.) on a single branch while tackling imbalanced biomedical datasets.

To validate that our design of the proposed deep transfer architecture supports further transfer learning on any other breast cancer biomedical dataset we performed experiments as illustrated in Table 6 using Breast-Histopathological-Images dataset. From the analysis it was validated that the VGGIN-Net architecture also supports transfer learning concept when tested on other breast cancer datasets.

5.2 Ablation Study

Ablation study has been conducted for the proposed VGGIN-Net architecture. In Table 7, we show experimental results on the 40X magnification factor to compare and contrast the use of block 3, 4 and 5 as the backbone features for our network. We inferred that feature extraction till block4 pool layer is an ideal combination. Another set of experiments were conducted to demonstrate the selection of naïve Inception block in the proposed architecture. Comparison with another variant of the Inception block as tabulated in Table 8 proves that the naïve inception block is apt for our model. Although, the dimensionality reduction block variant of the Inception module is less computationally expensive in comparison to the naïve inception block, the attained model performance obtained after combining the dimensionality reduction block to the proposed architecture is less in comparison to the naïve Inception block as validated by the experimental results illustrated in Table 8. The naïve Inception block was initially used with 64, 128, 32 filters for 1x1, 3x3, 5x5 conv layers respectively. In Table 9, we show experiments on using diverse widening factors (K) where each value of K indicates the multiple factor by which the filters are increased. It was found that K value as 1 is the ideal choice instead of K values of 2, 3, 5 and 10 in the naïve Inception block. Also, it helps to validate our choice of the number of filters besides keeping the computational complexity optimum since K=1 is the lowest possible considered value. More results in the supplementary file highlight the significance of data augmentation for enhancing the performance of the various deep networks including VGGIN-Net. Table 10 illustrates the experiments related to the proposed VGGIN-Net with and without data augmentation for different magnification factors for the BreakHis dataset. Experiments show that VGGIN-Net with data augmentation works significantly better in comparison to VGGIN-Net without Data Augmentation. The incorporation of data augmentation in the training pipeline helps reduce over-fitting by imparting the necessary regularization, allowing the models to learn continually across several epochs. For our case, we have applied all random transformations including random cropping of samples to a fixed crop size.

TABLE 2
PERFORMANCE EVALUATION OF VGG16, GOOGLENET, AND RESNET-50 WITH THE MODIFIED VGG16 ARCHITECTURE AND THE PROPOSED APPROACH ON BREAKHIS DATASET (I) 40X, (II) 100X, (III) 200X, (IV) 400X

Technique	40X				100X			
	Accuracy	F1	IBA	GMean	Accuracy	F1	IBA	GMean
VGG16 [19]	0.9294	0.87 / 0.95	0.79	0.89	0.9240	0.86 / 0.95	0.80	0.89
GoogLeNet [20]	0.8682	0.78 / 0.91	0.71	0.84	0.8674	0.78 / 0.91	0.72	0.85
ResNet50 [26]	0.9350	0.89 / 0.95	0.85	0.92	0.9381	0.89 / 0.96	0.86	0.93
Modified VGG16 w/ Single Dense Layer	0.9387	0.89 / 0.96	0.84	0.91	0.9522	0.92 / 0.97	0.90	0.95
Modified VGG16 w/ Inception Block w/ Single Dense Layer	**0.9628**	**0.93 / 0.97**	**0.93**	**0.96**	**0.9681**	**0.95 / 0.98**	**0.93**	**0.96**

Technique	200X				400X			
	Accuracy	F1	IBA	GMean	Accuracy	F1	IBA	GMean
VGG16 [19]	0.9119	0.86 / 0.94	0.83	0.91	0.8913	0.81 / 0.92	0.72	0.85
GoogLeNet [20]	0.8880	0.82 / 0.92	0.78	0.88	0.8668	0.77 / 0.91	0.69	0.83
ResNet50 [26]	0.9431	0.90 / 0.96	0.87	0.93	0.9221	0.87 / 0.94	0.81	0.90
Modified VGG16 w/ Single Dense Layer	0.9357	0.88 / 0.96	0.82	0.91	0.8893	0.82 / 0.92	0.77	0.88
Modified VGG16 w/ Inception Block w/ Single Dense Layer	**0.9651**	**0.88 / 0.96**	**0.80**	**0.89**	**0.9364**	**0.89 / 0.95**	**0.86**	**0.93**

https://github.com/SainiManisha/VGGIN-Net

TABLE 3
COMPARISON OF THE PROPOSED APPROACH WITH THE STATE-OF-THE-ART APPROACHES ON BREAKHIS DATASET BASED ON MEAN ACCURACY ACROSS DIFFERENT MAGNIFICATION FACTORS

Technique	40X	100X	200X	400X
Spanhol et al. [8]	0.8960 ± 0.0650	0.8500 ± 0.0480	0.8400 ± 0.0320	0.8080 ± 0.0310
Spanhol et al. [39]	0.8460 ± 0.0290	0.8480 ± 0.0420	0.8420 ± 0.0170	0.8160 ± 0.0370
Bayramoglu et al. [10]	0.8300 ± 0.0300	0.8310 ± 0.0350	0.8460 ± 0.0270	0.8210 ± 0.0440
Zhu et al. [40]	0.8570 ± 0.0190	0.8420 ± 0.0320	0.8490 ± 0.0220	0.8010 ± 0.0440
Gupta et al. [41]	0.8674 ± 0.0237	0.8856 ± 0.0273	0.9031 ± 0.0376	0.8831 ± 0.0301
Deniz et al. [17]	0.9096 ± 0.0159	0.9058 ± 0.0196	0.9137± 0.0172	0.9130 ± 0.0740
Song et al. [42]	0.9002 ± 0.0302	0.9120 ± 0.0440	0.8780 ± 0.0530	0.8740 ± 0.0720
Gupta et al. [18]	0.9471 ± 0.0088	0.9590 ± 0.0420	0.9676 ± 0.0109	0.8911 ± 0.0012
Ours	**0.9588 ± 0.0033**	**0.9657 ± 0.0087**	**0.9500 ± 0.0122**	**0.9315 ± 0.0034**
Ours (with fine tuning)	**0.9710 ± 0.0046**	**0.9667 ± 0.0022**	**0.9716 ± 0.0033**	**0.9368 ± 0.0053**

TABLE 4
PERFORMANCE EVALUATION OF THE PROPOSED APPROACH WITH UNDERSAMPLING AND OVERSAMPLING TECHNIQUES ON BREAKHIS DATASET (I) 40X, (II) 100X, (III) 200X, (IV) 400X

Sampling Technique	40X				100X			
	Accuracy	F1	IBA	GMean	Accuracy	F1	IBA	GMean
Undersampling	0.9591	0.93 / 0.97	0.89	0.94	0.9381	0.90 / 0.96	0.89	0.94
Oversampling	0.9406	0.90 / 0.96	0.89	0.94	0.9593	0.93 / 0.97	0.90	0.95
None	**0.9628**	**0.93 / 0.97**	**0.93**	**0.96**	**0.9681**	**0.95 / 0.98**	**0.93**	**0.96**

Sampling Technique	200X				400X			
	Accuracy	F1	IBA	GMean	Accuracy	F1	IBA	GMean
Undersampling	0.9540	0.91 / 0.97	0.86	0.92	0.9262	0.88 / 0.95	0.84	0.91
Oversampling	0.9669	0.94 / 0.98	0.92	0.96	0.9303	0.88 / 0.95	0.83	0.91
None	**0.9651**	**0.88 / 0.96**	**0.80**	**0.89**	**0.9364**	**0.89 / 0.95**	**0.86**	**0.93**

TABLE 5
PERFORMANCE EVALUATION OF BLOCK WISE FINE-TUNING ON PROPOSED VGGIN-NET FOR 40X, 100X, 200X AND 400X MAGNIFICATION FACTORS ON BREAKHIS DATASET

Fine Tuning	40X				100X			
	Accuracy	F1	IBA	GMean	Accuracy	F1	IBA	GMean
Complete Network	0.9666	0.99 / 0.97	0.89	0.94	**0.9753**	**0.96 / 0.98**	**0.96**	**0.98**
Block 2, 3, 4, Inception block	0.9777	0.96 / 0.98	0.95	0.97	0.8674	0.71 / 0.91	0.56	0.74
Block 3, 4, Inception block	**0.9851**	**0.97 / 0.99**	**0.96**	**0.98**	0.9646	0.94 / 0.98	0.89	0.94
Block4, Inception	0.9610	0.93 / 0.97	0.88	0.94	0.9700	0.95 / 0.98	0.92	0.96

https://github.com/SainiManisha/VGGIN-Net

block No Fine Tuning	0.9628	0.93 / 0.97	0.93	0.96	0.9752	0.95 / 0.98	0.93	0.97

Fine Tuning	200X				400X			
	Accuracy	F1	IBA	GMean	Accuracy	F1	IBA	GMean
Complete Network	**0.9688**	**0.95 / 0.98**	**0.92**	**0.96**	0.9077	0.86 / 0.93	0.85	0.93
Block 2, 3, 4, Inception block	0.9467	0.91 / 0.96	0.91	0.95	0.9323	0.89 / 0.95	0.83	0.91
Block 3, 4, Inception block	0.9651	0.94 / 0.98	0.89	0.94	0.9426	0.90 / 0.96	0.85	0.92
Block 4, Inception block	0.9614	0.93 / 0.97	0.92	0.96	**0.9528**	**0.92 / 0.97**	**0.91**	**0.95**
No Fine Tuning	0.9651	0.88 / 0.96	0.80	0.89	0.9364	0.89 / 0.95	0.86	0.93

TABLE 6
TRANSFER LEARNING OF PROPOSED VGGIN-NET ON BREAST HISTOPATHOLOGICAL DATASET WITH AND WITHOUT FINE-TUNING

Transfer Learning	Accuracy	F1	IBA	GMean
VGGIN-Net as Fixed Feature Extractor	0.8470	0.89 / 0.73	0.66	0.81
Fine Tuning the VGGIN-Net Inception Block	**0.8678**	**0.91 / 0.75**	**0.67**	**0.82**

TABLE 7
ANALYSIS OF FEATURES EXTRACTED FROM DIFFERENT BLOCKS OF VGG16 ARCHITECTURE TO FIND THE APPROPRIATE FEATURES IN THE PROPOSED ARCHITECTURE FOR 40X MAGNIFICATION FACTOR ON BREAKHIS DATASET

Technique	40X			
	Accuracy	F1	IBA	GMean
Proposed Network using block3_pool	0.9536	0.92 / 0.97	0.89	0.94
Proposed Network using block4_pool	**0.9628**	**0.93 / 0.97**	**0.93**	**0.96**
Proposed Network using block5_pool	0.9536	0.92 / 0.97	0.89	0.94

TABLE 8
ANALYSIS OF PROPOSED ARCHITECTURE WITH THE INCEPTION BLOCK AND DIMENSIONALITY REDUCTION INCEPTION BLOCK FOR 40X MAGNIFICATION FACTOR ON BREAKHIS DATASET

Technique	40X			
	Accuracy	F1	IBA	GMean
Proposed Network w/ Naïve Inception Block	**0.9628**	**0.93 / 0.97**	**0.93**	**0.96**
Proposed Network w/ Dimensionality Reduction Inception Block	0.9443	0.90 / 0.96	0.82	0.90

TABLE 9
ANALYSIS OF APPROPRIATE NUMBER OF FILTERS IN INCEPTION BLOCK TO BE USED IN THE PROPOSED ARCHITECTURE FOR 40X MAGNIFICATION FACTOR ON BREAKHIS DATASET

Widening Factor	40X			
	Accuracy	F1	IBA	GMean
k=1	**0.9628**	**0.93 / 0.97**	**0.93**	**0.96**
k=2	0.9443	0.90 / 0.96	0.83	0.91
k=5	0.9684	0.95 / 0.98	0.93	0.96
k=10	0.9684	0.94 / 0.98	0.90	0.95

TABLE 10
PROPOSED VGGIN-NET WITH AND WITHOUT DATA AUGMENTATION FOR 40X, 100X, 200X AND 400X MAGNIFICATION FACTORS ON BREAKHIS DATASET

Technique	40X				100X			
	Accuracy	F1	IBA	GMean	Accuracy	F1	IBA	GMean
VGGIN-Net w/ Data Augmentation	**0.9628**	**0.93 / 0.97**	**0.93**	**0.96**	**0.9681**	**0.95 / 0.98**	**0.93**	**0.96**

VGGIN-Net w/o Data Augmentation	0.9239	0.86 / 0.95	0.80	0.89	0.9134	0.85/ 0.94	0.78	0.88

Fine Tuning	200X				400X			
	Accuracy	F1	IBA	GMean	Accuracy	F1	IBA	GMean
VGGIN-Net w/ Data Augmentation	**0.9651**	**0.88 / 0.96**	**0.80**	**0.89**	**0.9364**	**0.89 / 0.95**	**0.86**	**0.93**
VGGIN-Net w/o Data Augmentation	0.9155	0.85 / 0.94	0.78	0.88	0.8852	0.79/ 0.92	0.68	0.82

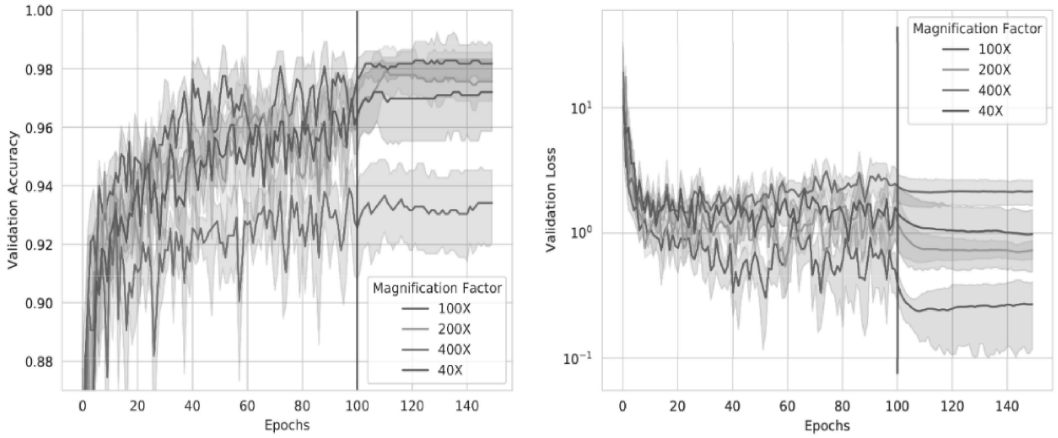


Figure 4. Validation accuracy and loss plot corresponding to the proposed architecture VGGIN-Net for different magnification factors (40X, 100X, 200X and 400X). Purple line indicate start of fine tuning.

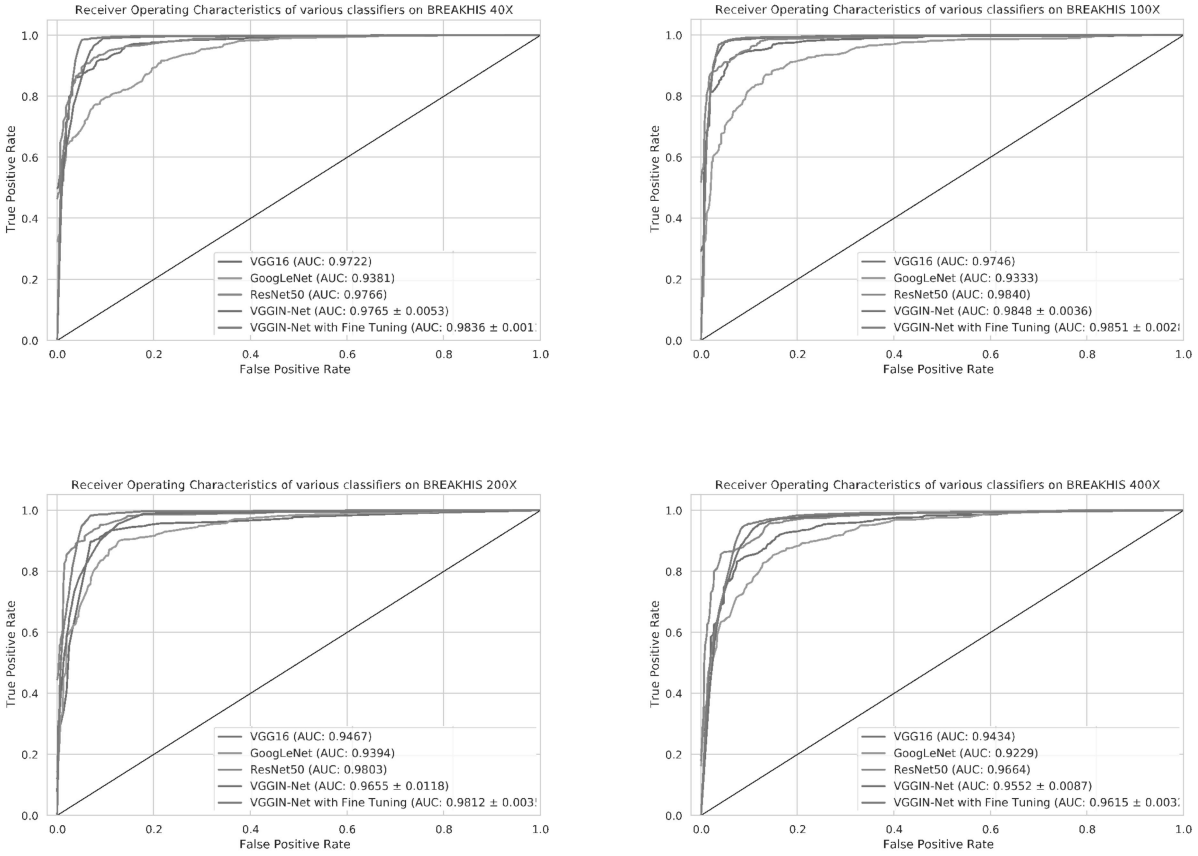


Figure 5. ROC curve comparison of proposed approach with state-of-the-art networks in case of (i) 40X, (ii) 100X, (iii) 200X and (iv) 400X

https://github.com/SainiManisha/VGGIN-Net

magnification factors for the BreakHis dataset

6 Conclusion and future scope

In this paper, a novel deep learning based network VGGIN-Net has been proposed using layers from the pre-trained deep network VGG-16 at lower level, and trainable
Inception module and dense layers at the higher level. The proposed transfer network has been compared with different state-of-the-art approaches on the basis of various performance evaluation metrics. It is validated from the experiments that VGGIN-Net designed to deal with the imbalanced breast cancer dataset and overall helps to improve robustness and generalizability of the approach. The proposed deep transfer network with fine tuning has achieved accuracies of 97.10%, 96.67%, 97.16% and 93.68% for the $40X$, $100X$, $200X$ and $400X$ magnification factors respectively, for the BreakHis dataset. The proposed network was able to classify both the minority and majority classes effectively. We also validated through experiments that the trained VGGIN-Net model supports transfer learning on other breast cancer datasets.
In future, we shall explore the use of skip connections inside deep neural nets. Also, we would look into constructing other deep network architectures for classification of multi-class imbalanced biomedical datasets. We would seek to create a hybrid approach based on techniques similar to DeTraC [43], combined with popular CNN pre-trained models to deal with multi-class imbalanced datasets.

References

[1] Lukong, Kiven Erique. "Understanding breast cancer–The long and winding road." BBA clinical 7 (2017): 64-77.

[2] Cheng, Heng-Da, Juan Shan, Wen Ju, Yanhui Guo, and Ling Zhang. "Automated breast cancer detection and classification using ultrasound images: A survey." Pattern recognition 43, no. 1 (2010): 299-317.

[3] Eklund, Anders, Paul Dufort, Daniel Forsberg, and Stephen M. LaConte. "Medical image processing on the GPU–Past, present and future." Medical image analysis 17, no. 8 (2013): 1073-1094.

[4] LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86, no. 11 (1998): 2278-2324.

[5] Yanai, Keiji, and Yoshiyuki Kawano. "Food image recognition using deep convolutional network with pre-training and fine-tuning." In 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 1-6. IEEE, 2015.

[6] Parvin, Hamid, Behrouz Minaei-Bidgoli, and Hamid Alinejad-Rokny. "A new imbalanced learning and dictions tree method for breast cancer diagnosis." Journal of Bionanoscience 7, no. 6 (2013): 673-678.

[7] Hamidinekoo, Azam, Erika Denton, Andrik Rampun, Kate Honnor, and Reyer Zwiggelaar. "Deep learning in mammography and breast histology, an overview and future trends." Medical image analysis 47 (2018): 45-67.

[8] Spanhol, Fabio Alexandre, Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte. "Breast cancer histopathological image classification using convolutional neural networks." In 2016 international joint conference on neural networks (IJCNN), pp. 2560-2567. IEEE, 2016.

[9] Feng, Yangqin, Lei Zhang, and Juan Mo. "Deep manifold preserving autoencoder for classifying breast cancer histopathological images." IEEE/ACM Transactions on Computational Biology and Bioinformatics (2018).

[10] Bayramoglu, Neslihan, Juho Kannala, and Janne Heikkilä. "Deep learning for magnification independent breast cancer histopathology image classification." In 2016 23rd International conference on pattern recognition (ICPR), pp. 2440-2445. IEEE, 2016.

[11] Bardou, Dalal, Kun Zhang, and Sayed Mohammad Ahmad. "Classification of breast cancer based on histology images using convolutional neural networks." IEEE Access 6 (2018): 24680-24693.

[12] Susan, Seba, and Amitesh Kumar. "SSOMaj-SMOTE-SSOMin: Three-step intelligent pruning of majority and minority samples for learning from imbalanced datasets." Applied Soft Computing 78 (2019): 141-149.

[13] Susan, Seba, and Amitesh Kumar. "Hybrid of Intelligent Minority Oversampling and PSO-Based Intelligent Majority Undersampling for Learning from Imbalanced Datasets." In International Conference on Intelligent Systems Design and Applications, pp. 760-769. Springer, Cham, 2018.

[14] Saini, Manisha, and Seba Susan. "Data Augmentation of Minority Class with Transfer Learning for Classification of Imbalanced Breast Cancer Dataset Using Inception-V3." In Iberian Conference on Pattern Recognition and Image Analysis, pp. 409-420. Springer, Cham, 2019.

[15] Saini, Manisha, and Seba Susan. "Comparison of Deep Learning, Data Augmentation and Bag of-Visual-Words for Classification of Imbalanced Image Datasets." In International Conference on Recent Trends in Image Processing and Pattern Recognition, pp. 561-571. Springer, Singapore, 2018.

[16] Rakhlin, Alexander, Alexey Shvets, Vladimir Iglovikov, and Alexandr A. Kalinin. "Deep convolutional neural networks for breast cancer histology image analysis." In International Conference Image Analysis and Recognition, pp. 737-744. Springer, Cham, 2018.

[17] Deniz, Erkan, Abdulkadir Şengür, Zehra Kadiroğlu, Yanhui Guo, Varun Bajaj, and Ümit Budak. "Transfer learning based histopathologic image classification for breast cancer detection." Health information science and systems 6, no. 1 (2018): 18.

[18] Gupta, Vibha, and Arnav Bhavsar. "Sequential modeling of deep features for breast cancer histopathological image classification." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 2254-2261. 2018.

[19] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[20] Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going deeper with convolutions." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1-9. 2015.

[21] Agostinelli, Forest, Matthew Hoffman, Peter Sadowski, and Pierre Baldi. "Learning activation functions to improve deep neural networks." arXiv preprint arXiv:1412.6830 (2014).

[22] Perdana, Anugrah Bintang, and Adhi Prahara. "Face Recognition Using Light-Convolutional Neural Networks Based On Modified Vgg16 Model." In 2019 International Conference of Computer Science and Information Technology (ICoSNIKOM), pp. 1-4. IEEE, 2019.

[23] Kumar, Abhinav, Sanjay Kumar Singh, Sonal Saxena, K. Lakshmanan, Arun Kumar Sangaiah, Himanshu Chauhan, Sameer Shrivastava, and Raj Kumar Singh. "Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer." Information Sciences 508 (2020): 405-421.

[24] Aravind, Krishnaswamy R., Purushothaman Raja, Rajendran

https://github.com/SainiManisha/VGGIN-Net

View publication stats

Ashiwin, and Konnaiyar V. Mukesh. "Disease classification in Solanum melongena using deep learning." Spanish Journal of Agricultural Research 17, no. 3 (2019): 0204.

[25] Baheti, Bhakti, Suhas Gajre, and Sanjay Talbar. "Detection of distracted driver using convolutional neural network." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1032-1038. 2018.

[26] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Identity mappings in deep residual networks." In European conference on computer vision, pp. 630-645. Springer, Cham, 2016.

[27] Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting." The journal of machine learning research 15, no. 1 (2014): 1929-1958.

[28] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." In Advances in neural information processing systems, pp. 1097-1105. 2012.

[29] Yosinski, Jason, Jeff Clune, Yoshua Bengio, and Hod Lipson. "How transferable are features in deep neural networks?." In Advances in neural information processing systems, pp. 3320-3328. 2014

[30] Shorten, Connor, and Taghi M. Khoshgoftaar. "A survey on image data augmentation for deep learning." Journal of Big Data 6, no. 1 (2019): 60.

[31] Howard, Andrew G. "Some improvements on deep convolutional neural network based image classification." arXiv preprint arXiv: 1312.5402 (2013).

[32] Sharma, Shallu, and Rajesh Mehra. "Conventional Machine Learning and Deep Learning Approach for Multi-Classification of Breast Cancer Histopathology Images—a Comparative Insight." Journal of Digital Imaging (2020): 1-23.

[33] Sharma, Shallu, and Rajesh Mehra. "Effect of layer-wise fine-tuning in magnification-dependent classification of breast cancer histopathological image." The Visual Computer 36, no. 9 (2020): 1755-1769.

[34] Kandel, Ibrahem, and Mauro Castelli. "How Deeply to Fine-Tune a Convolutional Neural Network: A Case Study Using a Histopathology Dataset." Applied Sciences 10, no. 10 (2020): 3359.

[35] Databases – Laboratório Visão Robótica e Imagem, 2019, https://web.inf.ufpr.br/vri/databases/. (Accessed 28 November 2019).

[36] Mooney, Paul. "Breast Histopathology Images." Kaggle. December 19, 2017. Accessed October 09, 2020. https://www.kaggle.com/paultimothymooney/breast-histopathology-images.

[37] Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, 2019.

[38] Sokolova, Marina, and Guy Lapalme. "A systematic analysis of performance measures for classification tasks." Information Processing & Management 45, no. 4 (2009): 427-437

[39] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte, "Deep features for breast cancer histopathological image classification," 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2017.

[40] Zhu, Chuang, Fangzhou Song, Ying Wang, Huihui Dong, Yao Guo, and Jun Liu. "Breast cancer histopathology image classification through assembling multiple compact CNNs." BMC medical informatics and decision making 19, no. 1 (2019): 198.

[41] Gupta, Vibha, and Arnav Bhavsar. "Breast cancer histopathological image classification: is magnification important?" In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 17-24. 2017.

[42] Song, Yang, Hang Chang, Heng Huang, and Weidong Cai. "Supervised intra-embedding of fisher vectors for histopathology image classification." In International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 99-106. Springer, Cham, 2017.

[43] Saini, Manisha, and Seba Susan. "Deep transfer with minority data augmentation for an imbalanced breast cancer dataset." Applied Soft Computing 97 (2020): 106759.

[44] Abbas, Asmaa, Mohammed M. Abdelsamea, and Mohamed Medhat Gaber. "Detrac: Transfer learning of class decomposed medical images in convolutional neural networks." IEEE Access 8 (2020): 74901-74913.

[45] Ding W, Huang D, Chen Z, Yu X, Lin W. Facial action recognition using very deep networks for highly imbalanced class distribution. In: 2017 Asia-Pacific signal and information processing association annual summit and conference (APSIPA ASC). 2017. p. 1368–72. https://doi.org/10.1109/APSIPA.2017.82822 46.

[46] Saini, Manisha, and Seba Susan. "Bag-of-Visual-Words codebook generation using deep features for effective classification of imbalanced multi-class image datasets." Multimedia Tools and Applications (2021): 1-27.



Ms. Manisha Saini is currently pursuing Ph.D. in Computer Science and Engineering Department from Delhi Technological University, Delhi, India and also working as Assistant Professor in the Computer Science Engineering Department at Faculty of Engineering and Technology, Manav Rachna International Institute of Research and Studies. Her research interests includes Computer Vision, Neural Networks, Machine Learning, and Deep Learning.



Dr. Seba Susan (M'17) is a Professor in the Department of Information Technology, Delhi Technological University, Delhi, India. She completed her Ph.D. from the Indian Institute of Technology (IIT) Delhi in 2014. Her current area of research is the development of soft computing tools for computer vision, speech and language processing.

https://github.com/SainiManisha/VGGIN-Net