SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(118th Edition)

OCTOBER 2022

BY: CENTRAL LIBRARY

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

PREFACE

This is the **One Hundred Eighteenth** 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 **October, 2022**. The arrangement of the contents is alphabetical. The full text of the article which is either subscribed by the university or available in the web is provided in this bulletin.

Central Library

CONTENTS

1. A Comparative Study of Demand Forecasting Models for a Multi-Channel Retail Company: A Novel Hybrid Machine Learning Approach, ***8.Arnab Mitra***, ***8.Arnav Jain***, ***8.Avinash Kishore*** and ***3.Pravin Kumar***, Mechanical, DTU

2. A deep neural network-based hybrid recommender system with user-user networks, ***8.1.Ayush Tanwar*** and ***3.Dinesh Kumar Vishwakarma***, CSE and IT, DTU

3. A Graph Convolution Network based Adaptive Cooperative Spectrum Sensing in Cognitive Radio Network, Dimpal Janu, ***6.1.Sandeep Kumar*** and ***6.1.Kuldeep Singh***, Electronics, DTU

4. A Novel Recommender System Using PSO, FCM and, Autoencoder, ***3.Om Prakash Verma*** and ***6.Nidhi Beniwal***, Electronics and IT, DTU

5. A review on modelling and techniques used for piezoelectric power generation from vibration of geo-structures, ***6.Nisha Kumari*** and ***3.Ashutosh Trivedi***, Civil, DTU

6. A Survey on Integration of Blockchain and IoT (BIoT): Open Issues, Challenges & Solution, ***8.Chirag Garg***, ***8.Deepak Kumar Mishra***, ***8.Dikshant Raj*** and ***3.Pawan Singh Mehra***, CSE, DTU

7. A survey on recent developments in diabetic retinopathy detection through integration of deep learning, ***6.Shalini Agarwal*** and ***3.Aruna Bhat***, CSE, DTU

8. A Systematic Review on Artificial Intelligence-Based Techniques for Diagnosis of Cardiovascular Arrhythmia Diseases: Challenges and Opportunities, ***6.Shikha Singhal*** and ***3.Manjeet Kumar***, Electronics, DTU

9. A Waste Segregation Bin, ***8.Saurav Kumar***, ***8.Vartika Kaushik*** and ***8.Annoy Roy***, Chemistyr and Production and Industrial Engineering, DTU

10. An Adaptive Intelligent Approach to Implement Energy Effi ciency Concept in Built Infrastructure, ***7.Srijan Singh***, ***7.Amisha Srivastava***, ***3.M. Rizwan*** and ***6.Astitva Kumar***, Electrical, DTU

11. An Effective Multimodal Biometric System Based on Textural Feature Descriptor, Neeru Bala, ***6.1.Rashmi Gupta*** and Anil Kumar, Electronics, DTU

12. An Inner Round Pipeline Architecture Hardware Core for AES, ***8.Archit Jain***, ***8.Divyanshu Jain***, ***8.Arpan Katiyar*** and ***3.Prof. Gurjit Kaur***, Electronics, DTU

13. An Optimized XGBoost approach for Predicting Progression of Hepatitis C using Hyperparameter Tuning and Feature Interaction Constraint, ***3.Kaustubh Ranjan Singh***, ***8.Ritik Gupta***, ***8.Ritik Kumar Kadian*** and ***8.Ritwik Singh***, Electronics, DTU

14. Analysis of Blended Concrete Cubes under Impact loading using ANSYS, ***6.Janhavi Singh*** and ***3.Shilpa Pal***, Civil, DTU

15. Analysis of Software Project Reports for Defect Prediction Using KNN, Rajni Jindal, ***3.Ruchika Malhotra*** and Abha Jain, CSE, DTU

16. ANALYZING AND PRIORITIZING RISKS INCLUDING CIRCULAR ECONOMY INITIATIVES IN SUPPLY CHAINS, ***6.Shweta Kumari*** and ***6.Akanksha Singh***, Economics, DTU

17. Appraisal of hydrochemistry and suitability assessment for water in an agriculture-dominated aquatic ecosystem of Rajasthan, India, ***7.Chitrakshi*** and ***3.Anil Kumar Haritash***, Environmental, DTU

18. APPROXIMATION BY DURRMEYER VARIANT OF CHENEY-SHARMA CHLODOVSKY OPERATORS, ***3.Chandra Prakash***, ***6.Durvesh Kumar Verma*** and ***3.Naokant Deo***, Mathematics, DTU

19. Assessment of sustainable maintenance performance of automobile garages in India, A. T. James, ***3.G. Kumar***, ***8.1.A. Pundhir***, ***8.1.S. Tiwari*** and ***8.R. Sharma***, DTU

20. Augmentation in performance and security of WSNs for IoT applications using feature selection and classification techniques, ***8.Rajiv Yadav***, ***3.Indu Sreedevi*** and ***3.Daya Gupta***, Electronics and CSE, DTU

21. Bifurcation Analysis of Current-mode controlled CUK converter, ***7.Avinash Kumar*** and ***3.Sudarshan K. Valluru***, Electronics, DTU

22. Bifurcation Analysis of Novel Self Operating Flyback Converter, ***7.Avinash Kumar*** and ***3.Sudarshan K. Valluru***, Electronics, DTU

23. Biomedical Implantable Antenna with Fractal Geometry, ***6.Nancy*** and ***3.Kaustubh Ranjan Singh***, Electronics, DTU

24. Blocking Probability Analysis for Energy Efficient Next Generation Green Cognitive Radio Networks, ***6.Akanksha Srivastava*** and ***3.Gurjit Kaur***, ECE, DTU

25. Board committees and financial performance: exploring the effects of gender diversity in the emerging economy of India, ***6.Shubham Singhania***, ***3.Jagvinder Singh*** and ***3.Deepti Aggrawal***, USME, DTU

26. Carrier Concentration Mediated Enhancement in Thermoelectric Performance of Various Polymorphs of Hafnium Oxide: A Plausible Material for High Temperature Thermoelectric Energy Harvesting Application, ***6.Rajesh Kumar***, Ramesh Kumar, ***6.Mukhtiyar Singh***, ***3.Deshraj Meena*** and Ankush Vij, Applied Physics, DTU

27. CERTAIN COEFFICIENT PROBLEMS OF S∗ e AND Ce, ***h*** AND ***8.1.NEHA VERMA***, Applied Mathematics, DTU

28. Character Segmentation on Indian Vehicle License Plate using K-means, OTSU and Niblack algorithm, ***8.Abhishek Jha***, ***8.Aman Singh***, ***6.Abdul Basit Ali*** and ***3.Rajni Jindal***, CSE, DTU

29. Coeffcient Functional and Bohr-Rogosinski Phenomenon for Analytic functions involving Semigroup Generators, ***7.Surya Giri*** and ***3.S. Sivaprasad Kumar***, Mathematics, DTU

30. Compact Dual-Band Zig Zag Shaped Implantable Antenna for Biomedical Devices, Rajiv Kumar Nehra and ***3.N S Raghava***, Electronics, DTU

31. Comparison of MPPT Algorithms Under Uniformly Varying Atmospheric Conditions, ***7.Chirag Kaushik***, ***3.Prof. Rachana Garg*** and ***3.Prof. Priya Mahajan***, Electrical, DTU

32. Deep Learning Models Performance Evaluations for Remote Sensed Image Classification, Abebaw Alem and ***3.Shailender Kumar***, CSE, DTU

33. Design and Analysis of LLC Resonant Converter and CCCV Topology for Battery Charging, ***7.Praphull Prakash Gupta***, ***3.Narendra Kumar*** and ***3.Uma Nangia***, Electrical, DTU

34. Design and Performance Analysis of High-Performance Low Power Voltage Mode Sense Amplifier for Static RAM, ***8.Divya DUTT***, ***3.Poornima MITTAL***, ***6.1.Bhawna RAWAT*** and Brijesh KUMAR, Electronics and Communication, DTU

35. Designs, Performance and Economic Feasibility of Domestic Solar Dryers, ***8.Shimpy***, ***8.Mahesh Kumar*** and ***3.Anil Kumar***, Mechanical, DTU

36. Dielectric frequency response at different voltage levels of mineral oil-based silica nanofluids, ***7.1.H.C. Verma*** and A. Baral, Electrical, DTU

37. Effect of Different Counter Electrodes on Power Conversion Efficiency of DSSCs, Geetam Richhariya, ***3.Anil Kumar***, Akash Kumar Shukla, K. N. Shukla and Bhim Charan Meikap, Mechanical and Environment, DTU

38. Emission performance assessment of in-use diesel-driven cars, ***6.Abhinav Pandey***, ***3.Rajeev Kumar Mishra*** and Govind Pandey, Environmental, DTU

39. Estimation of Daily Groundwater Table Using Backpropagation Neural Network Model by Assessing Training Algorithm, Varna Vishakar V, Ayush Jain, ***7.1.Zohaib Ahmed Khan*** and Ayush Kumar, Environment, DTU

40. Ethical Framework for Iot in People Analytics: Risks and Opportunities, ***6.Mr. Nishant Gaur*** and ***3.Dr. Vikas Gupta***, DSM, DTU

41. Experimental Analysis of Geopolymer Concrete:ASustainable and Economic Concrete Using the Cost Estimation Model, ***6.1.Manvendra Verma***, Kamal Upreti, Prashant Vats, Sandeep Singh, Prashant Singh, ***3.Nirendra Dev***, Durgesh Kumar Mishra and Basant Tiwari, Civil, DTU

42. Experimental study of the local scour around the two piers in the tandem arrangement using ultrasonic ranging transducers, ***6.Geeta Devi*** and ***3.Munendra Kumar***, Civil, DTU

43. Exploring Bitslicing Architectures for Enabling FHE-assisted Machine Learning, Soumik Sinha, Sayandeep Saha, Manaar Alam, ***8.1.Varun Agarwal***, Ayantika Chatterjee, Anoop Mishra, Deepak Khazanchi and Debdeep Mukhopadhyay, Electronics, DTU

44. Feasibility Study and Power Flow Management of DC Bus for Household Loads, ***6.Anupam***, ***3.Mayank Kumar*** and ***3.Suman Bhowmick***, Electrical, DTU

45. Functionalized Peptide-Based Nanoparticles for Targeted Cancer Nanotherapeutics: A State-of-the-Art Review, Ritika Sharma, Shikha Jyoti Borah, Bhawna, Sanjeev Kumar, Akanksha Gupta, ***3.Poonam Singh***, Vijay Kumar Goel, Ravinder Kumar and Vinod Kumar, Chemistry, DTU

46. Gene expression based inference of cancer drug sensitivity, Smriti Chawla, Anja Rockstroh, Melanie Lehman, Ellca Ratther, Atishay Jain, Anuneet Anand, ***6.ApoorvaGupta***, Namrata Bhattacharya, Sarita Poonia, Priyadarshini Rai, Nirjhar

Das, Angshul Majumdar, Jayadeva, Gaurav Ahuja, Brett G. Hollier, Colleen C. Nelson and Debarka Sengupta, Biotechnology, DTU

47. Hydrogen Generation from Biowaste & Its Application as A Fuel, ***8.Rishabh Sehgal***, ***6.Vijay Sharma*** and ***3.J. P.Kesari***, Applied Physics and Mechanical, DTU

48. Hydrogeochemical characterization of groundwater and their associated potential health risks, Anjali Singh, Ashwani Raju, Surendra Kumar Chandniha, ***6.Lipi Singh***, ***6.Inderjeet Tyagi***, Rama Rao Karri and Ajay Kumar, Environmental, DTU

49. Image dehazing using autoencoder convolutional neural network, Richa Singh, Ashwani Kumar Dubey and ***3.Rajiv Kapoor***, Electronics, DTU

50. IMPACT OF EMOTIONAL INTELLIGENCE ON JOB PERFORMANCE OF NURSES WITH MEDIATING EFFECT OF JOB SATISFACTION, Richa Chauhan, Vidhi Kaul and ***3.Nidhi Maheshwari***, DSM, DTU

51. Impact of Environmental and Social Reporting and Performance on Financial Performance of a Firm: An Indian Study, ***3.Varsha Sehgal***, ***3.Naval Garg*** and Jagvinder Singh, DSM, DTU

52. INTEGRAL MODIFICATION OF BETA-APOSTOL-GENOCCHI OPERATORS, ***6.Neha*** and ***3.Naokant Deo***, Mathematics, DTU

53. Integrating Blockchain and 5G: Advantages and Challenges of Decentralization, ***8.Dhruv Kumar***, ***8.Dhruv Kashyap***, ***6.Deepak Kumar Mishra*** and ***3.Pawan Singh Mehra***, CSE, DTU

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

55. Integration Schemes for Hybrid Generation Systems, Ashish Grover, Anita Khosla and ***3.Dheeraj Joshi***, Electrical, DTU

56. Intelligent-Routing Algorithm for wireless body area networks, Smita Sharma, V M Mishra and ***3.M M Tripathi***, Electrical, DTU

57. Measures of Information Reflect Memorization Patterns, ***8.Rachit Bansal***, Danish Pruthi and Yonatan Belinkov, Electrical, DTU

58. Minimization of torque ripples in PMSM drive using PI- resonant controller-basedmodel predictive control, ***6.Suryakant***, ***3.Mini Sreejeth***, ***3.Madhusudan Singh*** and ***6.1.Aakash Kumar Seth***, Electrical, DTU

59. Monitoring, modeling, and mapping of rail-induced noise at selected stations in megacity Delhi, K. Kumar, A. Bhartia and ***3.R. K. Mishra***, Environmental, DTU

60. Multimodal Human Eye Blink Recognition Using Z-score Based Thresholding and Weighted Features, Puneet Singh Lamba, Deepali Virmani, Manu S. Pillai and ***7.1.Gopal Chaudhary***, Electronics, DTU

61. Public Fund Care Tracking System based on Blockchain, ***8.Srishti Kumari***, ***8.Tushar Dixit***, ***3.Prem Prakash*** and ***6.Vaishali Sharma***, Electrical, DTU

62. Rapid Image Super Resolution, ***8.Keshav Gupta***, ***8.Divyansh Goel***, ***8.Divyani*** and ***3.Varun Sangwan***, Electronics, DTU

63. Recent advances in flavonoid-based nanocarriers as an emerging drug delivery approach for cancer chemotherapy, Saikat Dewanjeel, Pratik Chakraborty, Hiranmoy Bhattacharya, Sachin Kumar Singh, Kamal Dua, Abhijit Dey and ***6.1.Niraj Kumar Jha***, Biotechnology, DTU

64. Recycling Marine Plastic into Clothing Apparel via Global Collaborations, ***8.Akkshita Swain***, ***8.Tanisha Singhal*** and ***3.J. P. Kesari***, Physics and Mechanical, DTU

65. Retraction Note: Electrical power system transmission quality and power supplier micro grid control functional reliability, ***3.L. N. Das*** and ***8.1.Sanyam Gupta***, Mathematics, DTU

66. Second Generation Voltage Conveyer based Comparator and its application as Pulse Width Modulator, ***7.Abhinav Anand*** and ***3.Rajeshwari Pandey***, Electronics, DTU

67. Sharp Bounds of Fifth Coefficient and Hermitian-Toeplitz determinants for Sakaguchi Classes, ***7.Surya Giri*** and ***3.S. Sivaprasad Kumar***, Mathematics, DTU

68. Simulation of Photovoltaic based DC-DC Converter Topologies, Richa Khera, Anita Khosla, ***3.Dheeraj Joshi***, Electrical, DTU

69. The Inf luence of Board Factors and Gender Diversity on the ESG Disclosure Score: A Study on Indian Companies, Shikha Bhatia and ***6.Divya Marwaha***, DSM, DTU

70. Toeplitz Determinants in One and Higher Dimensions, ***7.Surya Giri*** and ***3.S. Sivaprasad Kumar***, Mathematics, DTU

71. TWO STAGE POWER CONVERSION FOR GRID CONNECTED PV USING CURRENT CONTROL TECHNIQUE, ***3.Mayank Kumar*** and ***3.Alka Singh***, Electrical, 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*

Operations Research Forum (2022) 3:58
https://doi.org/10.1007/s43069-022-00166-4

ORIGINAL RESEARCH

A Comparative Study of Demand Forecasting Models for a Multi-Channel Retail Company: A Novel Hybrid Machine Learning Approach

Arnab Mitra[1] · **Arnav Jain**[1] · **Avinash Kishore**[1] · **Pravin Kumar**[1]

Received: 26 March 2022 / Accepted: 22 August 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract

Demand forecasting has been a major concern of operational strategy to manage the inventory and optimize the customer satisfaction level. The researchers have proposed many conventional and advanced forecasting techniques, but no one leads to complete accuracy. Forecasting is equally important in manufacturing as well as retail companies. In this study, the performances of five regression techniques of machine learning, viz. random forest (RF), extreme gradient boosting (XGBoost), gradient boosting, adaptive boosting (AdaBoost), and artificial neural network (ANN) algorithms, are compared with a proposed hybrid (RF-XGBoost-LR) model for sales forecasting of a retail chain considering the various parameters of forecasting accuracy. The weekly sales data of a US-based retail company is considered in the analysis of the forecasts undertaking the attributes affecting the sale such as the temperature of the region and the size of the store. It is observed that the hybrid RF-XGBoost-LR outperformed the other models measured against various metrics of performance. This study may help the industry decision-maker to understand and improve the methods of forecasting.

Keywords Machine learning · Demand forecasting · Statistical methods · Random forest · XGBoost · AdaBoost · Gradient boosting

✉ Pravin Kumar
pravin.dce@gmail.com

Arnab Mitra
arnabmitra490@gmail.com

Arnav Jain
arnavjain17.4@gmail.com

Avinash Kishore
Avikish7@gmail.com

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

Published online: 27 September 2022

1 Introduction

In the present business world, organizations must improve their services in terms of efficiency, reliability, and availability to survive in the market. Sales forecasting and effective demand planning positively affect the performance of a supply chain [1]. The goal of demand planning is to develop a forecasting model that helps decision-makers in the areas of procurement, production, distribution, and sales. Forecasts serve as the basis for action plans carried out by various organizational units at different planning levels [2]. Managers usually predict sales based on their perceptions, intuitions, and experience. This may vary from person to person and it is very hard to continuously get reliable inputs from qualified and experienced managers. As a result, computer networks can assist in decision-making by estimating future sales. Machine learning (ML) can be utilized to create effective sales forecasting models utilizing the vast amount of data and related information [3].

ML techniques have become prevalent across various disciplines due to their ability to address the problems associated with increasingly large and complex datasets [4]. It involves complex algorithms to reveal meaningful patterns in large-scale and diverse datasets, which would be virtually impossible for a well-trained person [5]. Machine learning focuses on inductive inference inducing general models from specific empirical data [6]. In recent years, advancements in this field have been driven primarily by the creation of new algorithms as well as the ongoing burst of data at reduced computational costs [7].

ML methodsempowered by predictive analysis create enhanced customer engagement andforecast demands with better precision and accuracy in comparison to thetraditional demand forecasting methods [8, 9]. ML techniques can handle complicated correlations between so many causalelements having nonlinear relational demand patterns, thereby boosting retailchain performance [10]. The auto-regressive integrated moving average (ARIMA)and auto-regressive integrated moving average with exogenous variables (ARIMAX)approaches are the most often used predictive models for demand forecasting.Recently developed ML algorithms, such as artificial neural networks (ANN), supportvector machine (SVM), and regression trees, have already outperformed thetraditional methods [11]. The main objectives of the study are as follows:

- To explore the machine learning models used for forecasting.
- To compare the select machine learning models and hybrid models for sales forecasting of a US-based retail company.

In this study, several ML models are compared for retail demand forecasting. These models include random forecast (RF), artificial neural network (ANN), gradient boosting (GB), adaptive boosting (AdaBoost), and extreme gradient boosting (XGBoost), and the performance of these models was compared with the proposed hybrid model of the RF, XGBoost, and linear regression (LR). To make an accurate

comparison of the said models, various performance metrics, namely, mean squared error (MSE), R^2 score, and mean absolute error (MAE), were considered. The advantages and limitations of the employed methodologies as well as future options for performance enhancement are explored. The historical sales data of a leading US-based multinational retail company is considered for the forecasting analysis. The company has a large number of retail stores across the globe, specializing in a long range of products fulfilling the day-to-day demands of consumers. The sales data used for forecasting is related to various stores of the company spread across the USA.

The rest parts of the paper are arranged as follows: Sect. 2 represents the literature review, Sect. 3 discusses the research methodology, Sect. 4 represents the case study and result discussion, and Sect. 5 concludes the study.

2 Literature Review

Many advances and changes have been observed in the global business world during the past couple of decades. Some major factors leading to business-related uncertainties include partner activities, consumer behaviour, rival behaviour, evolving technology, and new product development [12]. Because of these uncertainties, the market is becoming complex and competitive and needs contemporary supply chain management [13]. Precise forecasting is essential for the success of supply chains [14]. On the other hand, various endogenous factors concerning the collection and application of field data can make the forecasting techniques extremely difficult and the external factors can also have a detrimental impact on forecast accuracy [15]. Effective demand forecasting can save more than 7% on the annual operating expenditures of a business [16]. Either qualitative or quantitative techniques can be employed for demand forecasting [17]. Executives' consensus, Delphi technique, historical analogies, and market research are used as qualitative demand forecasting. These techniques heavily rely on domain experts' subjective evaluations and lack decision models, which are data-driven. Quantitative methods, such as regression and time-series analytics, tend to be more systematic and dependable [18]. Regression methods, in particular, are concerned with determining the causal relationships between the independent and the dependent variables [19].

The notion of industry 4.0, ML, and artificial intelligence (AI) as an innovative framework is now being applied to supply chain analytics [20, 21]. Strategic planning, ad-hoc reporting, and end-user computation are common in business intelligence and analytics that aid in robust performance evaluation and management [22, 23]. Descriptive analytics is a technique that deals with happenings in the past, while diagnostic and predictive analytics deals with the happenings to be predicted in the future. For mitigating harmful impacts, the module prescriptive analytics comes in the application [24, 25].

2.1 Machine Learning Methods

ML methods have garnered considerable interest from researchers and practitioners in demand forecasting [26]. But, when it comes to the context of supply chain management, these methods have not been investigated properly. Although algorithms for ML are complex, they offer a variety of distinct and flexible demand forecasting models [27]. The effectiveness of different statistical and ML methods is heavily debated, making it difficult to draw gross generalizations regarding their efficacy [28]. Under varied circumstances, each model class may outperform the others [29]. Some of the most prominent ML methods are discussed in brief in the following paragraphs which have been compared with the proposed hybrid model.

Random Forest Punia et al. [30] introduced a hybrid forecasting method that is based on long short-term memory (LSTM) and random forest (RF). This first model utilizes LSTM to map the temporal characteristics of the data and then random forest is used to model the residuals of the LSTM network. The random forest section of the network is of vital significance as it provides a substantial edge in forecasting accuracy due to its ability to predict sudden changes due to holidays, promotions etc.

XGBoost Kang et al. [31] used the XGBoost hybrid model for tourism and trend prediction. They introduced location attributes and the time-lag effect of network search data to propose the hybrid model. The findings suggest that the spatiotemporal XGBoost composite model outperforms single forecasting approaches. There are several modifiable parameters in the XGBoost algorithm, including general, promotion, and learning objective parameters. They adopted the tree model to concentrate on the nonlinear interactions between the Baidu index and the number of tourists. Using the general parameters, the promotion parameters were changed according to the model chosen.

Gradient Boosting Machine Xenochristou et al. [32] measured the influence of the spatial scale on water demand forecasting. Multiple models were trained on UK daily consumption records for different aggregations of consumptions. Three different levels of spatial aggregation were created using properties' postcodes. A gradient boosting model for training on each of the configurations and prediction for water consumption was made for 1 day in the future. The results implied that the amount of spatial aggregation had a substantial influence on forecasting accuracy and errors can be minimized by utilizing additional explanatory variables.

AdaBoost Walker and Jiang [33] observed that the forecasts using AdaBoost are more accurate and reliable than those derived via a more traditional logistic regression method. They have analysed the importance of each predictor in the AdaBoost model to better understand the relative contribution of each factor to the overall predicted outcome. They observed that AdaBoost models are not easily interpretable as regression model coefficients.

ANN Jahangir et al. [34] used rough artificial neural network (R-ANN) approach to forecast plug-in electric vehicles travel behaviour (PEVs-TB) and PEV load on the distribution network. R-ANNs can increase the accuracy of forecasting findings due to their capacity to analyse phenomena with high uncertainty. Furthermore, two training methods are used in this paper—conventional error back propagation (CEBP) and Levenberg–Marquardt—which are specified using first- and second-order derivatives, respectively. In PEVs-TB, the results demonstrated that the Levenberg–Marquardt approach is more accurate.

It is observed that various authors advocated the varying level of performance of the different models of forecasting which vary on a case-to-case basis. It is very difficult to say which the best performing model is. In this study, RF, ANN, GB, XGBoost, AdaBoost, and hybrid models are tested and compared with each other for the sales forecasting of a US-based retail company.

The hybrid network has been developed by the separate training of the random forest and the XGBoost model on 67% of the data. Then a new dataset was created by generating the prediction from both the models for all the data points. This new dataset was passed as input to train the linear regression model and it predicted the final sales values. Some of the most common machine learning models used in forecasting are summarized in Table 1.

2.2 Measurement of Forecasting Accuracy

Chicco et al. [65] used healthcare information to forecast rates of obesity. R^2 was observed to be more accurate and complete than symmetric mean absolute percentage error (SMAPE). The value of R^2 becomes high if the analysis accurately predicts the majority of ground truth entities for every ground truth category taking into account their dispersion. The accuracy of various machine learning/deep learning models is compared using MSE, MAE, mean absolute percentage error (MAPE), root mean squared error (RMSE), and R^2 metrics [64]. Tsoumakas [3] examined the present state of ML algorithms for forecasting food-purchasing habits. It covers essential design concerns for forecasting food sales, such as the temporal granularity of sales figures, the intake factors to use for predicting sales, and the depiction of the sales evaluation function. It also looks at machine learning algorithms for predicting food sales and important measures like MAE, MSE, RMSE, and MASE, for evaluating forecasting accuracy. Ala'raj et al. [66] used Covid infection data to model and forecast Covid-19 outbreaks. They utilized a modified SEIRD (Susceptible, Exposed, Infectious, Recovered, and Dead) dynamic model and ARIMA model for prediction. The model prediction accuracy was estimated by using 5 metrics: AE, MSE, MLSE (maximum likelihood sequence estimation), normalized MAE, and normalized MSE. Ramos et al. [67] examined the efficiency of phase space models and ARIMA regressors as a tool for predictions of retail sales of five different types of women's footwear: boots, booties, flats, sandals, and shoes. RMSE, MAE, and MAPE were used to evaluate the ARIMA model.

Table 1 Applications of machine learning models in forecasting

Forecasting model	Paper	Description/applications of the forecasting models
Random forest	Islam and Amin [35]	Forecasting utilizing distributed random forest (DRF) and gradient boosting machine (GBM) is examined in this study, which results in extremely simple potential backorder decision scenarios
	Mueller [36]	The predictive capabilities of random forest (RF) and conditional inference forest (CF) are measured. This study explores forecasting using RF which is developed using classification and regression trees (CART) and conditional inference tree (CIT)
	Punia et al. [30]	The research introduced a new forecasting approach that integrates deep learning (DL) with RF and long short-term memory (LSTM) networks. The model has many practical implications for retail managers for a more accurate estimation of complex demand patterns
	Li et al. [37]	The ensemble empirical mode decomposition random forest (EEMDRF) model is introduced for estimating daily electricity usage in conventional organizations. This data is then simulated and predicted via the random forest model
	Rao et al. [38]	A two-stage syncretic cost-sensitive random forest (SCSRF) model is proposed to assess the borrowers' credit risk. It is useful in improving classification and forecasting accuracy
	Ni et al. [39]	RF improved grey ideal value approximation (IGIVA), complementary ensemble empirical mode decomposition (CEEMD), particle swarm optimization algorithm based on dynamic inertia factor (DIFPSO), and backpropagation neural network (BPNN) are combined to develop a hybrid forecasting model. The hybrid model is compared with other models in analysing a PV power plant dataset
XGBoost	Wang et al. [40]	A new load forecasting approach based on SampEn variational mode decomposition (SVMD) and XGBoost is described for industrial customers in this paper. To mitigate detrimental impacts, an adaptive SVMD approach is developed and the XGBoost model is utilized in forecasting industrial customers' short-term electricity demand
	Kang, et al. [31]	A spatiotemporal XGBoost composite model for tourist trend prediction is presented by using geographical features and the time-lag of network search data. The spatiotemporal XGBoost hybrid model outperforms single forecasting approaches, like XGBoost or ARIMA
	Yun et al. [41]	A hybrid genetic algorithm-based XGBoost forecasting methodology is developed. This study empirically demonstrates the significance of feature engineering in forecasting the prices of stocks
	Wang et al. [42]	To overcome the challenges of predicting rental bike demand on an hourly basis, a data mining technique is used. It also investigates a feature filtering strategy to remove non-predictive variables and ranks features based on their prediction performance. Several models are considered in the analysis, namely, XGBoost, LR, and SVM

Table 1 (continued)

Forecasting model	Paper	Description/applications of the forecasting models
	Jabeur et al. [43]	For monthly streamflow forecasting, a hybrid model called Gaussian mixed model (GMM-XGBoost) is developed in which GMM was used to cluster streamflow into multiple groups and XGBoost is used for the prediction
	Osman et al. [44]	Six ML methods have been compared to forecast the gold price; two of them include XGBoost and CatBoost. XGBoost performed the best with the highest R^2 score
	Shi et al. [45]	XGBoost, artificial neural network (ANN), and support vector machine (SVM) are the three ML models tested for forecasting the groundwater levels; the models analysed 11 months of previously observed rainfall, temperature, and evaporation data. The R^2 score of the XGBoost model was the highest
AdaBoost	Walker and Jiang [33]	This study outlines a reproducible adaptive boosting (AdaBoost) model implementation which forecasts the demand-driven acquisition. AdaBoost model's forecasting power is compared to that of a more typical model of logistic regression, and the findings reveal that the AdaBoost model outperforms the others
	Zhou and Lai [46]	Three AdaBoost models are investigated for bankruptcy estimating with missing data by merging them with imputation approaches, and each model is tested on two datasets. For enterprises with or without missing financial data, the AdaBoost M1. Decision Trees (ABM1.DT) algorithm with the KNN approach is reliable in predicting bankruptcy
	Barrow and Crone [47]	This research assesses empirical forecast performance and follows a robust and reliable experimental methodology. The existing AdaBoost.RT and AdaBoost.R2 algorithms can be enhanced by modifying the meta-parameters and the developed AdaBoost.BC
	Wang et al. [48]	To improve generalization ability and reduce ensemble cost, a novel sparse AdaBoost is developed as an amalgamation model. The proposed AdaBoost-ESN (echo state network) technique performs well. AdaBoost-ESN reduces fitting error by 38.25% and forecasting error by 54.96%, respectively
	Sidhu et al. [49]	To anticipate irrigation in rice that is drip-irrigated, ML models such as the linear model, SVM, decision tree, AdaBoost, and RF were utilized. For all of the parameters, AdaBoost offered the best results, with the smallest error value and the biggest R^2 number
	Busari and Lim [50]	AdaBoost-GRU was proposed, wherein the GRU (gated recurrent unit) system was developed, wrapped in sklearn, and then boosted using AdaBoost Regressor. Utilizing daily crude oil prices, the developed framework's forecasting power was compared against AdaBoost-LSTM (long short-term memory). The AdaBoost-GRU ensemble model outperforms AdaBoost-LSTM in terms of delivering the least errors

Table 1 (continued)

Forecasting model	Paper	Description/applications of the forecasting models
	Huang et al. [51]	The proposed method BP AdaBoost (backpropagation AdaBoost) was compared in this research against a single BPNN (backpropagation neural network) model and a simple weighted average BPNN model. The mean value of RMSE was reduced by 8.23%
	Sun et al. [52]	In this study, the ability of the AdaBoost model for bankruptcy forecasting, specifically in the construction industry, has been demonstrated. The predictive powers of AdaBoost and several other algorithms were compared. Based on the experimental results, it was observed that AdaBoost had a more predictive ability as compared to the other models
	Heo and Yang [53]	AdaBoost and LSTM are combined to develop a hybrid ML model. MAPE (mean absolute percentage error) and DS (day-specific) assessment criteria comparison results demonstrate that the suggested approach's out-of-sample forecasting performance is superior to the benchmarks for all the time series
Gradient boosting	Sharma et al. [54]	The block-wise methodology is used to create five separate prediction models: GB, GB-PCA, ANN-CG, and ANN-CG-PCA. By merging the forecasts from the four different models, three hybrid forecasts are developed. The results reveal that the combined model performs better than the separate models, with a MAPE improvement of roughly 15%
	Deng et al. [55]	A hybrid model combining GBDT (gradient boosting decision tree) and DE (differential evolution) is designed to detect insider trading utilizing data from pertinent indexes. The proposed technique is best in terms of recognition accuracy when using a 90-day time window
	Xenochristou et al. [32]	A gradient boosting machine is used to forecast water demand 1 to 7 days in advance. For a reduction in group size from 600 to 5 households, the findings indicate an exponential deterioration in forecasting accuracy from a mean absolute percentage error (MAPE) of 3.2 to 17%
	Yoon [56]	This study examined ML algorithms to project real GDP growth using data from the IMF and the BOJ. The gradient boosting model has a higher forecasting accuracy than the RF model, according to the findings
	Gu et al. [57]	This study introduced a hybrid strategy focused on GBDT, correlation analysis, and EWT for forecasting the London Metal Exchange Nickel settlement price. The developed model was evaluated on three sets of financial data, and the findings demonstrate that when compared to the GBDT, it optimized the TIC output by 16.33%

Table 1 (continued)

Forecasting model	Paper	Description/applications of the forecasting models
	Nie et al. [58]	A new energy consumption prediction model is developed to simulate and predict building electrical energy use. The gradient boosting regression tree approach is used in the proposed model to more precisely estimate energy consumption data. The given prediction model is more accurate and faster to compute
Artificial neural network	Güven and Şimşir [59]	ML techniques have been used for sales forecasting in the retail garment business. ANN and SVM were both trained and evaluated. The ANN model without colour detail outperforms other models by delivering lower RMSE for a larger number of datasets
	Yucesan et al. [60]	Several models such as LR, auto-regressive integrated moving average (ARIMA), ANN, and exponential smoothing were tested while also being evaluated. In addition to this, hybrid models, namely, ARIMA-ANN and ARIMA-LR, were created and utilized in this study. The hybrid ARIMA-ANN achieved a MAPE of 0.49 which was the best out of all the models
	Fanoodi et al. [61]	ANN and ARIMA algorithms are used to predict the demand for blood platelet and reduce the supply uncertainties. It was observed that the ANN and the ARIMA models were far more accurate as compared to the baseline model which was used at Zahedan Blood Transfusion Center to calculate demand volatility
	Jahangir et al. [34]	In this study, an AI-based forward feed and recurrent ANN with an LM training module have been developed to predict the PEV-TB. With the help of the designed model, the net economic loss was reduced by approximately $16 per year per PEV when compared with conventional methods
	Jebaraj et al. [62]	This study showed that the ANN-based model can be utilized for forecasting coal demand in India as this model gave less MPE and higher R^2 scores for most cases
	Zhao and Yue [63]	The multi-objective linear weighted approach and ANN model are utilized to forecast the security in the supply management of the forest industry
	Loureiro et al. [64]	This study looks into how a DL model can be used in the estimation of sales in the fashion industry. The sales prediction obtained via the DL technique has also been compared with shallow techniques such as decision trees, RF, and SVM. DNN outperformed the other methods in part of the performance metrics considered

3 Methodology

In this study, the performances of XGBoost, RF, ANN, gradient boosting, AdaBoost, and the proposed hybrid framework (RF-XGBoost-LR) are compared using several performance metrics, namely, MAE, MSE, and R^2 score (coefficient of determination). XGBoost, RF, gradient boosting, and AdaBoost are ensemble techniques built on top of decision trees, while ANN is a deep learning technique. The framework of a decision tree (DT) comprises a root node (topmost node), internal nodes, and leaf nodes (end nodes). Simple principles are used in DT algorithms to branch out of the root node, passing through internal nodes and eventually ending in the leaves [68]. In this work, Python 3.7.12 was utilized. For data handling, Pandas version 1.1.5 and Numpy version 1.19.5 were used. For model training, XGBoost version 0.90 and Scikit-learn library version 1.0.2 were used.

3.1 Proposed Framework

In this study, a hybrid model of RF-XGBoost-LR is proposed and its performance is compared with other individual models.

Bagging vs Boosting The primary distinction between the approach of bagging and boosting methods is that the former decreases the variance in prediction by generating the additional data for training from the dataset using combinations with repetitions to produce multi-sets of the original data, while the later adjusts the weight of an observation based on the last classification by iteration. Unlike the bagging method, wherein a uniform selection of each sample is made to build a training dataset, the boosting algorithm's likelihood of choosing a particular sample is unequal. Misclassified or inaccurately calculated samples are more likely to be selected when they carry a higher weight. As a result, each new model may focus on samples that have incorrectly been classified by earlier models [69].

Random Forest RF is an ensemble technique in which the results of many regression trees are combined to generate a single prediction. The primary premise is bagging, in which a sample of training data is selected at random and fitted into a regression tree [70]. This randomly selected sample is termed a bootstrap sample and it is chosen with replacement, meaning any previously chosen data point can be chosen again. A bootstrap sample can be made by choosing N data points randomly from the dataset and then substituting them with the data points present in the dataset. There is a 1/N chance of any data point being chosen.

RF is a combination of decision tree estimators $\{h(X, \Theta\mathrm{k}),\ \mathrm{k} = 1, 2, ...\}$, in which every decision tree is calculated by utilizing the outputs of a random vector $\{\Theta\mathrm{k}\}$, which is independently sampled and evenly distributed among all the decision trees present in the forest.

Once the training is complete, the result of the entire set of decision trees on sample X′ is averaged to generate predictions as shown in Eq. (1).

$$\widehat{f} = \frac{1}{k} \sum_{i=1}^{k} h(X', \Theta k) \tag{1}$$

where $\widehat{f}$ is the final prediction and k is the number of decision trees.

XGBoost An abbreviation for 'extreme gradient boosting' is XGBoost with potential improvements upon gradient boosting. XGBoost enhances the performance and is capable of solving problems of real-world scale while making use of a minimum number of resources [71]. XGBoost is a parallel tree model built upon the gradient boosting model. It utilizes the tree ensemble method, which is made up of a series of CART. Although XGBoost consists of various unique characteristics, second-order Taylor expansion and embedded normalization algorithms appear to be similar to GBDT [72]. XGBoost models have the advantage of scaling effectively for different scenarios while requiring fewer resources than existing prediction models. Within XGBoost, parallel and distributed computation speeds up the model learning and allows for more rapid model exploration.

Hybrid (RF-XGBoost-LR) Model RF makes parallel decision trees which help in reducing the overfitting problem. Accuracy improves as a result of the reduction in variance. In RF, individually separate decision trees are used for each of the multiple copies of original training data. Despite its widespread popularity, random forest suffers from conceptual and practical shortcomings. Random forest adaptive learning is inherently poor in terms of minimizing training error. In particular, each tree is learned autonomously. Complementary information from other trees is not fully realized in this kind of training [73]. It results in a reduction in model performance. XGBoost combines multiple weak learners in a sequential method, which iteratively improves model performance.

XGBoost is a boosting technique. It takes advantage of parallel processing and runs the model on several CPU cores. However, it is affected by the renowned overfitting problem in boosting, which also impacts multiple additive regression trees (MARTs). This challenge arises when there are few trees accessible early in the iteration process; as a result, all of the trees impact significantly the model [74]. Overfitting of training data degrades the model's generalization capabilities, resulting in unreliable performance when applied to novel measurements. High variance and low bias estimators are common manifestations of overfitting. The extra complexity may, and frequently does, aid the model's performance on a set of training data, but it inhibits future point prediction [75]. Overfitting gives an overly optimistic impression of prediction results in new data drawn from the underlying population [76].

Hybrid models are combinations of two or more single models of machine learning or soft computing to achieve higher flexibility along with higher capability in contrast to a single model. One of the two entities, prediction, and optimization of

the prediction are often present in a hybrid model for higher accuracy. Mainly, there are two key reasons to develop a hybrid model:

(i) To eliminate the risk of an unfortunate prediction of a single forecast in some specific conditions.
(ii) To improve upon the performance of the independent models.

A hybrid model is designed to reap the advantages and overcome the shortcomings of the individual models involved [77]. In this research, a hybrid ML model has been proposed within which random forest regressor which is a bagging technique and XGBoost regressor which is a boosting technique have been combined.

A hybrid model has been developed to overcome the shortcomings of both the models, i.e. RF and XGBoost, as shown in Fig. 1. The random forest model addresses the overfitting problem inherent to XGBoost as it can decrease the model variance without increasing the model bias. This implies that the overfitting problem may be observed in the forecast of a single regression tree, but it can be eliminated in the average forecast of multiple regression trees. Random forest model is poor in terms of reducing the training error as multiple regression trees are trained autonomously. XGBoost addresses this shortcoming of random forest by sequentially training decision trees.

In the proposed framework, RF and XGBoost models are trained separately and predictions of both the models are used as input into an LR model. The LR model processes the final output.

The LR equation can be defined by Eq. (2):

$$Y = \beta + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \tag{2}$$

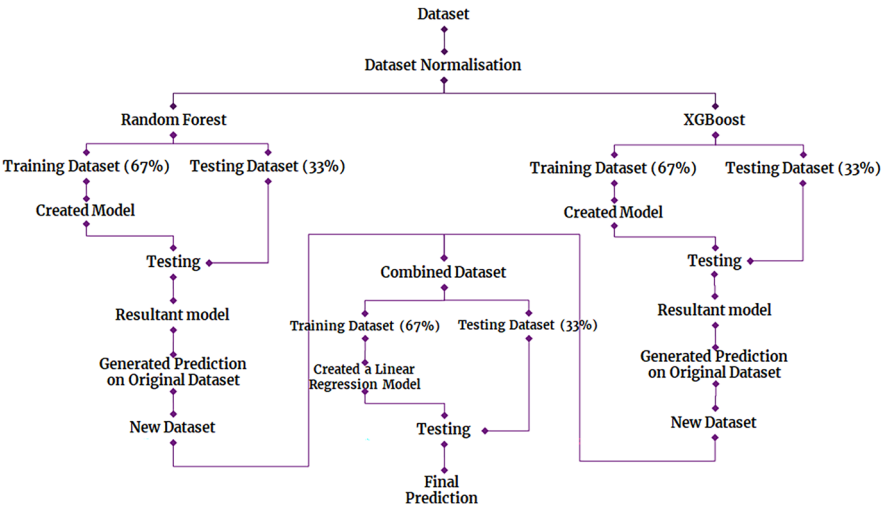


Fig. 1 Hybrid model flowchart

where the final prediction is represented by Y, and the predictions from RF and XGBoost are represented by x_1 and x_2. The y-intercept is represented using β. The coefficients and error terms are represented by β_1, β_2, and ε respectively. The value of β_1 is −0.05671126, β_2 is 1.05822592, and β is −4.4399934967759985e-05.

The Python source codes for all the four ML forecasting models and hybrid models are shown in the Appendix.

4 Case Study and Result Discussions

The case company operates as a merchandiser of consumer products. The international segment manages supercentres, supermarkets, hypermarkets, warehouse clubs, and cash and carries outside of the USA. The company was founded in 1945 and is headquartered in Bentonville, Arkansas. Among the largest retailers in the world, based in the USA, the company experiences revenue gain year over year. It operates grocery stores, supermarkets, hypermarkets, department stores, and discount stores offering commodities at the lowest prices, the strategy which defines it, in more than 25 countries across the globe. Fuel, gift cards, banking services, and other associated products such as money orders, prepaid cards, and wire transfers are all available through the company.

According to statistics, grocery prices were reduced by an average of 10–15% in markets where the company entered. It has a wide product range, which makes it a tough competitor among other companies in the same segment. Products offered range from electronics and offices, movies, music, and books to jewellery, baby products, and furniture for pharmacies. It is capable of lowering grocery prices by another significant margin during promotional periods. The strong market power over the supplier and competitors allows them to sell the products at the lowest prices and helps them compete in the market.

In this research, the data of a retail company known to keep up with the demands of customers by offering a wide range of products at one stop has been used. The sales data of the company spans different regions in the USA. The data consists of weekly sales for all the 45 stores and 99 departments over 3 years. The data consists of different attributes of the store and geographic-specific information, namely store number, size of the store, department of the store, date mentioning the week, region's average temperature, fuel price in the region, CPI (consumer price index), unemployment rate, and holiday week.

Normalization is a pre-processing step that plays a crucial role in machine learning. Normalizing aids in decreasing the learning time when the datasets are too large. Min–Max normalization transforms the original dataset into the desired interval using a linear transformation. This technique has the advantage of preserving all relations between the data points. Min–Max normalization is given by Eq. (3).

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{3}$$

For proportionate scaling of the data, the Min–Max scale was used at the beginning of the analysis keeping the minimum and maximum values as 0 and 1 respectively.

4.1 Performance Parameters

For the comparison of the forecasting models, mean absolute error (MAE), mean squared error (MSE), and R^2 value are used as discussed in the following subsections.

Mean Absolute Error It is an average measure of errors in a set of predictions. Since it is absolute, it ignores the positivity or negativity of the error and all individual errors are equally weighted. The calculation of MAE is straightforward as shown in Eq. (4). To get the 'total error', the absolute values of the errors are summed up and divided by the total number of observations [78].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{4}$$

where γ_i is the true value, $\hat{\gamma}_i$ is the prediction value, and n is the number of observations.

Mean Squared Error It is also an average measurement of the errors in a set of predictions. The squares of each error are added together and then averaged as shown in Eq. (5). This ensures that all errors are equal in weight and that the direction of the error is irrelevant. Since it is a quadratic function, it will always reach global minima.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{5}$$

where γ_i is the true value, $\hat{\gamma}_i$ is the prediction value, and n is the number of observations.

R^2 Score It is also known as the coefficient of determination which expresses the amount of variance in the dependent variable explained by a model [79]. R^2 score is used to evaluate the scattered data about a fitted regression line. Higher R^2 values for similar datasets represent smaller differences between the predicted data and the true data. It measures the relationship between predicted and true data on a scale of 0–1. For example, an R^2 value of 0.8 indicates that the variation of the independent variable explains 80% of the variance of the dependent variable being analysed. It is given by Eq. (6).

$$R^2 = 1 - \frac{SS_{res}}{SS_{total}} = \frac{\sum\left(y_i - \hat{y}_i\right)^2}{\sum\left(y_i - \mu\right)} \tag{6}$$

where SS_{res} is the sum of squares of residuals, SS_{total} is the total sum of squares, γ_i is the true value, $\widehat{\gamma}_i$ is the prediction value, and μ is the mean.

R^2 mainly shows whether the said model provides the goodness of fit for the observed values. It was also necessary to understand errors for which metrics namely MSE and MAE were utilized. Mean squared error (MSE)'s purpose was to put more effort into outliers. Due to its square, it weighs large errors more heavily than small ones. Mean absolute error (MAE) is used when measuring the prediction in the same unit as the original series. MAE provides information about how much an average error is expected from the forecast on average. Using the above performance parameters (MSE, MAE, R^2), all the different models incorporated in this study are compared as shown in Table 2.

In the RF model, the number of estimators was kept at 175 and the maximum depth was kept at 28. In the gradient boosting model, the number of estimators was kept at 125 and the maximum depth was kept at 25. In the AdaBoost model, a decision tree with a depth of 25 as a base estimator was used. In ANN, the model has kept 5-layer deep followed by an output layer. The number of neurons in each layer was 10, 12, 24, 12, and 10. The activation function for each layer was taken as 'relu' and 'adam' optimizer. The batch size was taken as 256 and the model was trained for 500 epochs. In the XGBoost model, the number of estimators was kept at 150 and the maximum depth was given as 25. In the RF-XGBoost-LR model, at first, the RF and the XGBoost models were used and their predictions were used as input to an LR model.

In the hyper-parameter optimization phase of the machine learning model, determining the most optimal configuration parameters of the ML optimization methods is challenging. As a result, using random values within the effective range of relevant ML algorithm parameters may result in enhanced optimization outcomes. The output from RF and XGBoost models is being passed as input to an LR model. The LR model was used to output the final predictions because of its simplicity. If we use a complex model like gradient boosting in our final layer, the hybrid model would be prone to overfitting. The hybrid model leads to overcoming the shortcomings of both the RF and XGBoost models.

Wolpert [80] proposed stacking (also known as stacked generalization), which is an ensemble of well-performing models for their capabilities. Stacking uses a single

Table 2 Performance of the forecasting models

S. no	Model	MSE	MAE	R^2
1	Random forest	6.92e-05	0.0028	0.9351
2	XGBoost	4.82e-05	0.0024	0.9547
3	Gradient boosting	9.42e-05	0.0032	0.9116
4	AdaBoost	7.38e-05	0.0027	0.9308
5	Artificial neural network	6.45e-04	0.0129	0.3958
6	RF-XGBoost-LR (hybrid)	**4.79e-05**	**0.0024**	**0.9551**

The bold values shows the performance of the proposed model

model to combine the different predictions from multiple models. Stacked models provide the best results by using a wide range of algorithms in the first layer of design, as different algorithms identify trends and patterns differently in training data, and merging both models results in a more accurate and reliable output.

Various performance measures were utilized to compare the performance of all the models. Holistically, based on three metrics, the proposed forecasting method is found to outperform the other benchmarking methods with an R^2 score of 0.9551, MAE of 0.0024, and MSE of 4.7932e-05.

Figure 2 shows the comparison of the week-wise sale of all the stores and departments against the sales predicted by the hybrid (RF-XGBoost-LR) model. It is observed from Table 2 that the hybrid networks can better forecast the sales as compared to the other models since they can map the trend of the actual sales most accurately.

4.2 Academic Implications

The proposed hybrid model can be utilized to enhance supply chain-related studies and be applied to extend research work on demand forecasting. The robust performance of the proposed framework augments its utility. Retailers, wholesalers, and other industries can use it to their benefit. It is however essential to have adequate domain knowledge for it to be tailored for various applications in different industries. Products in different sectors may have distinct properties that may be retrieved and put into the forecasting framework to fully increase their performance. This makes industry-specific customization to be a potential research topic.

4.3 Managerial Implications

The findings of the study show that the proposed hybrid model improves the forecasting accuracy up to a large extent compared to other individual machine learning

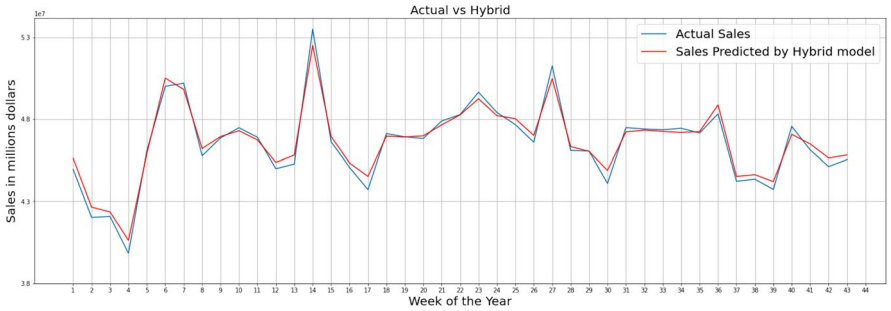


Fig. 2 Comparison between actual sales and forecasting using the hybrid model

models. Both the models random forest and XGBoost jointly overcome the problems of overfitting and training errors in linear regression analysis of the data, and hence, the forecast values are very close to the actual values of sales. Thus, the proposed model helps the industry decision-makers in more accurate forecasting, which leads to the formulation of a better marketing strategy, increasing stock turnover, optimizing capacity building, lowering supply chain costs, and improving customer happiness. An accurate demand forecasting method can improve the supply chain performance by eliminating the bullwhip effect and proper inventory management.

5 Conclusion

In this study, a hybrid model of ML has been proposed combining XGBoost, RF, and LR, for real-time analysis of sales data. Sales data of the retail company with various attributes are trained to introduce a newer more advanced model of veracity. To address the shortcomings of both the RF and XGBoost models, a hybrid model is proposed. At first, the dataset was normalized and then trained and tested separately in RF and XGBoost models. The predictions from these models were assimilated to create a new dataset, which was used as input into the LR model to generate the final predictions.

It is observed that by combining the XGBoost model with the RF model, the dataset improves the accuracy due to reduced variance and enhanced robustness to outliers, which results in improved predictive ability and less vulnerability to overfitting. Three metrics were used in this study, which are MAE, MSE, and R^2 scores. The results suggest that the proposed hybrid model RF-XGBoost-LR (MAE = 0.0024, MSE = 4.7932e-05, and R^2 score = 0.9551) has better performance than the other models, namely RF, ANN, gradient boosting, AdaBoost, and XGBoost. *R*-squared score infers that the model explains 95.51% of the data and variables incorporated.

A precise demand forecasting in an integrated commercial planning environment can be utilized to optimize capacity building, schedule labour management, inventory, supply chain management etc. In the proposed hybrid model, random forest helps to overcome the overfitting problem of XGBoost, while XGBoost is used to reduce the error by training the decision trees sequentially. Forecasting using the proposed model may improve stock availability and enhance stock allocation.

With all the implications, the proposed hybrid model has some limitations in terms of the requirement of the bid size of the training data, decision integration etc. As the size of training datasets expands, machine learning algorithms become more effective.

Appendix

Python Source Code

```
#Importing the necessary libraries
from sklearn.ensemble import RandomForestRegressor,GradientBoostingRegressor,AdaBoostRegressor
from sklearn.tree import DecisionTreeRegressor
import sklearn
import xgboost
from sklearn.model_selection import train_test_split,GridSearchCV
from sklearn.metrics import mean_absolute_error,mean_squared_error,r2_score
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import MinMaxScaler
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LinearRegression


#Reading The Dataset
train=pd.read_csv('train.csv')
stores=pd.read_csv('stores.csv')
features=pd.read_csv('features.csv')

#Combining The Data
df = train.merge(features,how="left",on=['Store','Date','IsHoliday']).merge(stores,how='left',on='Store')
df['IsHoliday'] = df['IsHoliday'].astype("int")
df.drop(['MarkDown1','MarkDown2','MarkDown3','MarkDown4','MarkDown5'],axis=1,inplace=True)

#Scaling The Data
X_data,Y_data=np.array(df.drop(['Date','Weekly_Sales'],axis=1).values),np.array(df['Weekly_Sales'].values)
scaler_x = MinMaxScaler()
xscale=scaler_x.fit_transform(X_data)
Y_data=np.reshape(Y_data, (-1,1))
scaler_y = MinMaxScaler()
yscale=scaler_y.fit_transform(Y_data)

#Train Test Split
x_train, x_test, y_train, y_test = train_test_split(xscale,yscale, test_size=0.33, random_state=42)


#Random Forest Model
rf=RandomForestRegressor(n_estimators=175,max_features='auto',max_depth=28)
rf.fit(x_train,y_train)
y_pred_rf=rf.predict(x_test)
print('R2_Score',r2_score(y_test,y_pred_rf))
print('MAE',mean_absolute_error(y_test,y_pred_rf))
print('MSE',mean_squared_error(y_test,y_pred_rf))
y_rf=rf.predict(xscale)
prediction_frame = pd.DataFrame(data=list(y_rf),columns=['RandomForest'])
prediction_frame['Actual']=yscale

#XGBoost Model
xgb=xgboost.XGBRegressor(max_depth=28,n_estimators=150)
xgb.fit(x_train,y_train)
y_pred_xgb=xgb.predict(x_test)
print('R2_Score',r2_score(y_test,y_pred_xgb))
print('MAE',mean_absolute_error(y_test,y_pred_xgb))
print('MSE',mean_squared_error(y_test,y_pred_xgb))
y_xgb=xgb.predict(xscale)
prediction_frame['XGB']=list(y_xgb)


# Creating A DataFrame From All The Predictions
prediction_frame.to_csv('Prediction.csv',index=False)
data=pd.read_csv('Prediction.csv')

# Creating Train And Test Data From The Predictions DataFrame
x_train, x_test, y_train, y_test = train_test_split(data[['RandomForest','XGB']],data['Actual'], test_size=0.33,random_state=42)

#Hybrid Model Using Random Forest and XGBoost Predictions
lr=LinearRegression()
lr.fit(x_train,y_train)
y_pred_lr=lr.predict(x_test)
y_pred_lr=y_pred_lr
print('R2_Score',r2_score(y_test,y_pred_lr))
print('MAE',mean_absolute_error(y_test,y_pred_lr))
print('MSE',mean_squared_error(y_test,y_pred_lr))
```

Data Availability The datasets generated during and/or analysed during the current study are available in the Kaggle.com repository (https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting/data).

Declarations

Competing Interests The authors declare no competing interests.

References

1. Kantasa-Ard A, Nouiri M, Bekrar A, Ait el Cadi A, Sallez Y (2021) Machine learning for demand forecasting in the physical internet: a case study of agricultural products in Thailand. Int J Prod Res 59(24):7491–7515
2. Haberleitner H, Meyr H, Taudes A (2010) Implementation of a demand planning system using advance order information. Int J Prod Econ 128(2):518–526
3. Tsoumakas G (2019) A survey of machine learning techniques for food sales prediction. Artif Intell Rev 52(1):441–447
4. Wilson ZT, Sahinidis NV (2017) The ALAMO approach to machine learning. Comput Chem Eng 106:785–795
5. Goecks J, Jalili V, Heiser LM, Gray JW (2020) How machine learning will transform biomedicine. Cell 181(1):92–101
6. Hüllermeier E (2015) Does machine learning need fuzzy logic? Fuzzy Sets Syst 281:292–299
7. Holzinger A (2016) Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Inf 3(2):119–131
8. Bohanec M, Borštnar MK, Robnik-Šikonja M (2017) Explaining machine learning models in sales predictions. Expert Syst Appl 71:416–428
9. Chase CW Jr (2016) Machine learning is changing demand forecasting. J Bus Forecast 35(4):43
10. Ampazis N (2015) Forecasting demand in supply chain using machine learning algorithms. Int J Artif Life Res (IJALR) 5(1):56–73
11. Smolak K, Kasieczka B, Fialkiewicz W, Rohm W, Siła-Nowicka K, Kopańczyk K (2020) Applying human mobility and water consumption data for short-term water demand forecasting using classical and machine learning models. Urban Water J 17(1):32–42
12. Sillanpää V, Liesiö J (2018) Forecasting replenishment orders in retail: value of modelling low and intermittent consumer demand with distributions. Int J Prod Res 56(12):4168–4185
13. Mohammed A (2020) Towards 'gresilient' supply chain management: a quantitative study. Resour Conserv Recycl 155:104641
14. Oliva R, Watson N (2009) Managing functional biases in organizational forecasts: a case study of consensus forecasting in supply chain planning. Prod Oper Manag 18(2):138–151
15. Van der Laan E, van Dalen J, Rohrmoser M, Simpson R (2016) Demand forecasting and order planning for humanitarian logistics: an empirical assessment. J Oper Manag 45:114–122
16. Van Wassenhove LN, Pedraza Martinez AJ (2012) Using OR to adapt supply chain management best practices to humanitarian logistics. Int Trans Oper Res 19(1–2):307–322
17. Holt CC (2004) Forecasting seasonals and trends by exponentially weighted moving averages. Int J Forecast 20(1):5–10
18. Maia ALS, de Carvalho FDA (2011) Holt's exponential smoothing and neural network models for forecasting interval-valued time series. Int J Forecast 27(3):740–759
19. Wang CH, Chen JY (2019) Demand forecasting and financial estimation considering the interactive dynamics of semiconductor supply-chain companies. Comput Ind Eng 138:106104
20. Jacobs FR, Chase RB, Lummus RR (2014) Operations and supply chain management (pp 533–535). New York, NY: McGraw-Hill/Irwin
21. Stevenson WJ, Hojati M, Cao J (2014) Operations management (p. 182). Chicago-USA: McGraw-Hill Education
22. Lu WM, Wang WK, Lee HL (2013) The relationship between corporate social responsibility and corporate performance: evidence from the US semiconductor industry. Int J Prod Res 51(19):5683–5695

23. Wang CH, Chen YW (2016) Combining balanced scorecard with data envelopment analysis to conduct performance diagnosis for Taiwanese LED manufacturers. Int J Prod Res 54(17):5169–5181
24. Addo-Tenkorang R, Helo PT (2016) Big data applications in operations/supply-chain management: a literature review. Comput Ind Eng 101:528–543
25. Hazen BT, Skipper JB, Ezell JD, Boone CA (2016) Big data and predictive analytics for supply chain sustainability: a theory-driven research agenda. Comput Ind Eng 101:592–598
26. Abolghasemi M, Hyndman RJ, Tarr G, Bergmeir C (2019) Machine learning applications in time series hierarchical forecasting. arXiv preprint arXiv:1912.00370
27. Abolghasemi M, Beh E, Tarr G, Gerlach R (2020) Demand forecasting in supply chain: the impact of demand volatility in the presence of promotion. Comput Ind Eng 142:106380
28. Aye GC, Balcilar M, Gupta R, Majumdar A (2015) Forecasting aggregate retail sales: the case of South Africa. Int J Prod Econ 160:66–79
29. Ahmed NK, Atiya AF, Gayar NE, El-Shishiny H (2010) An empirical comparison of machine learning models for time series forecasting. Economet Rev 29(5–6):594–621
30. Punia S, Nikolopoulos K, Singh SP, Madaan JK, Litsiou K (2020) Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. Int J Prod Res 58(16):4964–4979
31. Kang J, Guo X, Fang L, Wang X, Fan Z (2021) Integration of Internet search data to predict tourism trends using spatial-temporal XGBoost composite model. Int J Geogr Inf Sci 36(2):236–252
32. Xenochristou M, Hutton C, Hofman J, Kapelan Z (2020) Water demand forecasting accuracy and influencing factors at different spatial scales using a gradient boosting machine. Water Resources Res 56(8):e2019WR026304
33. Walker KW, Jiang Z (2019) Application of adaptive boosting (AdaBoost) in demand-driven acquisition (DDA) prediction: a machine-learning approach. J Acad Librariansh 45(3):203–212
34. Jahangir H, Tayarani H, Ahmadian A, Golkar MA, Miret J, Tayarani M, Gao HO (2019) Charging demand of plug-in electric vehicles: forecasting travel behaviour based on a novel rough artificial neural network approach. J Clean Prod 229:1029–1044
35. Islam S, Amin SH (2020) Prediction of probable backorder scenarios in the supply chain using distributed random forest and gradient boosting machine learning techniques. J Big Data 7(1):1–22
36. Mueller SQ (2020) Pre-and within-season attendance forecasting in Major League Baseball: a random forest approach. Appl Econ 52(41):4512–4528
37. Li C, Tao Y, Ao W, Yang S, Bai Y (2018) Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition. Energy 165:1220–1227
38. Rao C, Liu M, Goh M, Wen J (2020) 2-stage modified random forest model for credit risk assessment of P2P network lending to "Three Rurals" borrowers. Appl Soft Comput 95:106570
39. Ni L, Wang D, Wu J, Wang Y, Tao Y, Zhang J, Liu J (2020) Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. J Hydrol 586:124901
40. Wang Y, Sun S, Chen X, Zeng X, Kong Y, Chen J, Wang T (2021) Short-term load forecasting of industrial customers based on SVMD and XGBoost. Int J Electr Power Energy Syst 129:106830
41. Yun KK, Yoon SW, Won D (2021) Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process. Expert Syst Appl 186:115716
42. Wang Z, Hong T, Piette MA (2020) Building thermal load prediction through shallow machine learning and deep learning. Appl Energy 263:114683
43. Jabeur SB, Mefteh-Wali S, Viviani JL (2021) Forecasting gold price with the XGBoost algorithm and SHAP interaction values. Ann Operations Res 1–21
44. Osman AIA, Ahmed AN, Chow MF, Huang YF, El-Shafie A (2021) Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. Ain Shams Eng J 12(2):1545–1556
45. Shi R, Xu X, Li J, Li Y (2021) Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization. Appl Soft Comput 109:107538
46. Zhou L, Lai KK (2017) AdaBoost models for corporate bankruptcy prediction with missing data. Comput Econ 50(1):69–94
47. Barrow DK, Crone SF (2016) A comparison of AdaBoost algorithms for time series forecast combination. Int J Forecast 32(4):1103–1119
48. Wang L, Lv SX, Zeng YR (2018) Effective sparse adaboost method with ESN and FOA for industrial electricity consumption forecasting in China. Energy 155:1013–1031
49. Sidhu RK, Kumar R, Rana PS (2020) Machine learning based crop water demand forecasting using minimum climatological data. Multimed Tools Appl 79(19):13109–13124

50. Busari GA, Lim DH (2021) Crude oil price prediction: a comparison between AdaBoost-LSTM and AdaBoost-GRU for improving forecasting performance. Comput Chem Eng 155:107513
51. Huang H, Zhang Z, Song F (2021) An ensemble-learning-based method for short-term water demand forecasting. Water Resour Manage 35(6):1757–1773
52. Sun S, Wei Y, Wang S (2018) AdaBoost-LSTM ensemble learning for financial time series forecasting. Int Conf Comput Sci (pp 590–597). Springer, Cham
53. Heo J, Yang JY (2014) AdaBoost based bankruptcy forecasting of Korean construction companies. Appl Soft Comput 24:494–499
54. Sharma V, Cali Ü, Sardana B, Kuzlu M, Banga D, Pipattanasomporn M (2021) Data-driven short-term natural gas demand forecasting with machine learning techniques. J Petrol Sci Eng 206:108979
55. Deng S, Wang C, Wang M, Sun Z (2019) A gradient boosting decision tree approach for insider trading identification: an empirical model evaluation of China stock market. Appl Soft Comput 83:105652
56. Yoon J (2021) Forecasting of real GDP growth using machine learning models: gradient boosting and random forest approach. Comput Econ 57(1):247–265
57. Gu Q, Chang Y, Xiong N, Chen L (2021) Forecasting nickel futures price based on the empirical wavelet transform and gradient boosting decision trees. Appl Soft Comput 109:107472
58. Nie P, Roccotelli M, Fanti MP, Ming Z, Li Z (2021) Prediction of home energy consumption based on gradient boosting regression tree. Energy Rep 7:1246–1255
59. Güven İ, Şimşir F (2020) Demand forecasting with color parameter in retail apparel industry using artificial neural networks (ANN) and support vector machines (SVM) methods. Comput Ind Eng 147:106678
60. Yucesan M, Gul M, Celik E (2018) A multi-method patient arrival forecasting outline for hospital emergency departments. Int J Healthcare Manage 13(Sup1):283–295
61. Fanoodi B, Malmir B, Jahantigh FF (2019) Reducing demand uncertainty in the platelet supply chain through artificial neural networks and ARIMA models. Comput Biol Med 113:103415
62. Jebaraj S, Iniyan S, Goic R (2011) Forecasting of coal consumption using an artificial neural network and comparison with various forecasting techniques. Energy Sources Part A Recov Util Environ Effects 33(14):1305–1316
63. Zhao X, Yue S (2021) Analysing and forecasting the security in supply-demand management of Chinese forestry enterprises by linear weighted method and artificial neural network. Enterprise Inf Syst 15(9):1280–1297
64. Loureiro AL, Miguéis VL, da Silva LF (2018) Exploring the use of deep neural networks for sales forecasting in fashion retail. Decis Support Syst 114:81–93
65. Chicco D, Warrens MJ, Jurman G (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 7:e623
66. Ala'raj M, Majdalawieh M, Nizamuddin N (2021) Modeling and forecasting of COVID-19 using a hybrid dynamic model based on SEIRD with ARIMA corrections. Infect Dis Model 6:98–111
67. Ramos P, Santos N, Rebelo R (2015) Performance of state space and ARIMA models for consumer retail sales forecasting. Robot Comput Integr Manuf 34:151–163
68. Parsa AB, Movahedi A, Taghipour H, Derrible S, Mohammadian AK (2020) Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. Accid Anal Prev 136:105405
69. Zhang Y, Haghani A (2015) A gradient boosting method to improve travel time prediction. Transport Res Part C Emerg Technol 58:308–324
70. Lahouar A, Slama JBH (2015) Day-ahead load forecast using random forest and expert input selection. Energy Convers Manage 103:1040–1051
71. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system, in proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16). San Francisco, CA, 785–794
72. Kaplan UE, Dagasan Y, Topal E (2021) Mineral grade estimation using gradient boosting regression trees. Int J Min Reclam Environ 35(10):728–742
73. Ren S, Cao X, Wei Y, Sun J (2015) Global refinement of random forest. Proc IEEE Conf Comput Vision Pattern Recogn 723–730
74. Samat A, Li E, Wang W, Liu S, Lin C, Abuduwaili J (2020) Meta-XGBoost for hyperspectral image classification using extended MSER-guided morphological profiles. Remote Sens 12(12):1973

75. Jabbar H, Khan RZ (2015) Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). Computer Sci Commun Instrument Devices 70
76. Steyerberg EW (2019) Overfitting and optimism in prediction models. Clin Predict Models (pp 95–112). Springer, Cham.)
77. Ardabili S, Mosavi A, Várkonyi-Kóczy AR (2019) Advances in machine learning modeling reviewing hybrid and ensemble methods. In International Conference on Global Research and Education (pp 215–227). Springer, Cham
78. Wang Z, Bovik AC (2009) Mean squared error: love it or leave it? A new look at signal fidelity measures. IEEE Signal Process Mag 26(1):98–117
79. Valbuena R, Hernando A, Manzanera JA, Görgens EB, Almeida DR, Silva CA, García-Abril A (2019) Evaluating observed versus predicted forest biomass: R-squared, index of agreement or maximal information coefficient? Eur J Remote Sens 52(1):345–358
80. Wolpert DH (1992) Stacked generalization. Neural Netw 5(2):241–259

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

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

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-022-13936-3

A deep neural network-based hybrid recommender system with user-user networks

Ayush Tanwar[1] · **Dinesh Kumar Vishwakarma**[2]

Received: 21 December 2021 / Revised: 13 April 2022 / Accepted: 12 September 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

In today's digital age, choosing the right product, web page, news article, or even a research paper like this one from an extensive number of options is one of the most tedious tasks. The resolution to this problem is using a recommender system(RS), which helps you choose the suitable item according to your profile. In this research, We present a novel deep neural network based hybrid recommender system that addresses the lacunas of traditional Collaborative Filtering (CF) and current hybrid systems while also delivering higher accuracy in recommendations. Due to insufficient training data, CF recommender systems suffer from low accuracy, linear latent factor, and cold-start problem. To overcome these problems, we employ a Deep neural network-based approach which uses user and item vectors to encapsulate users' and items' data to train on High dimensionality non-linear data to provide more accurate recommendations. User-user networks are employed to provide a better collaboration and synergy facet to our model. In our approach, Combining user-user networks with Deep neural networks yields higher predictive accuracy and better running time than other state-of-art methods. Extensive experimentation on publicly available Flixster and MovieLens Datasets concludes that our technique outperforms current premier methods by achieving improvement of 19% in RMSE, 9.2% in MAE and 4.1% in F1 Score.

Keywords Recommender system · User networks · Deep learning system · Hybrid recommender systems · Collaborative filtering

✉ Dinesh Kumar Vishwakarma
dinesh@dtu.ac.in

Ayush Tanwar
ayushtw@gmail.com

[1] Department of Computer Science and Engineering, Delhi Technological University, New Delhi, 110042, Delhi, India

[2] Department of Information Technology, Delhi Technological University, New Delhi, 110042, Delhi, India

Published online: 12 October 2022

1 Introduction

In today's high-tech world, we interact with data all the time. A large volume of data is available to users in a diversity of formats, and the rate of production of data is too fast for users to process efficiently; As a result, the users suffer from Information overload. Smartphones and other smart devices create colossal amounts of data from various mediums like video streaming, social media, sharing of multimedia, hosting services, instant messengers, emails, etc. All this increases the difficulty of finding suitable content on the world wide web and makes the user more susceptible to making an error while choosing. This error can be detrimental to the platform with which the user interacts. For example, if the user selects a movie that is not suitable for his or her taste, the user is more likely not to complete the movie and dislike the movie and the streaming service altogether. What the users need is an intelligent system like a Recommender system (RS) [21], which will help them choose the right content according to their penchant.

A recommender system (or RS) is an intelligent application-based system, which is designed to deal with information overload problem and to provide reasonable suggestions on items to its users [16]. The RS's ability to offer efficient recommendations for relevant items is beneficial; this ability has helped many social media giants like Facebook, YouTube, and Netflix accumulate huge revenues. All these factors have made Recommender Systems a searing and crucial area of research. In simplistic words, what RS does is to make a personalized list of ordered items for users to pick from; order of the items is calculated from various sources of data obtained from the users like their likes, dislikes, interests, demographics, etc. These personalized RS can provide recommendations in various domains like news [39], ecommerce [30], music [18] and movies [16]. RS can provide support to a plethora of tools and services on the big platforms. Some of the most crucial functions of RS are increasing engagement, increasing sales, and improving user satisfaction and user experience.

1.1 Collaborative filtering (CF)

CF RS [34] gives suggestions to the target user by anticipating the target user's rating for a specific target item based on the ratings of the user's top-n alike users. The similarity in this technique is calculated using a rating matrix. A rating matrix is a $n \; x \; m$ matrix which consist of ratings of *'n'* items given by *'m'* users. CF systems suffer from cold start [19] and sparsity problems. The Cold start problem is the situation when the system cannot draw any presumptions for users or items due to insufficient information. New systems, users, or items are most affected by this problem. Our approach alleviates this problem by using domain-based attributes in our neural network.

1.2 Content based approaches

Content based approaches [32] use domain knowledge to calculate recommendations. All the feature representations are hand-engineered. Therefore, plenty of domain knowledge is required for a content-based recommender system. The model is as good as the knowledge with which the feature representations were created. Moreover, content-based approaches can give recommendations using only the existing preferences of the user. In the future, if users' preferences expand or change, the feature representations will need to be changed. To overcome this, we have combined CF techniques with content-based techniques to create a hybrid RS [19].

In this paper, we present a new Deep Neural network-based hybrid Recommender System with user-user networks, which builds on the capabilities offered by conventional approaches like collaborative filtering (CF) [34] and content-based approaches like Ontology [32].

1.3 Why deep neural network-based hybrid recommender system with user-user networks?

Recommendations organically use word-of-mouth approach and similar people like similar things approach. An individual is more likely to buy an item that was recommended or liked by a connection than an item that was not. User-user networks in our model use this principle and recommend items based on the user's connections. Employing user-user networks in the recommendation process provides an excellent collaboration and synergy factor. Rather than just calculating the similarity based on attributes between two users, we calculate similarity according to the shortest distance between the two users in the user-user network. The closer two people are, the higher the likelihood that the people belong to the same socio-economic sphere. Moreover, a shorter distance between two people also indicates a higher probability that the two people have interacted with each other or with similar items.

Using only the user-user network will eventually lead to improved collaboration and synergy, but it will also lead to poor personalized recommendations. Because with improved collaboration and synergy, most people will get recommendations for only popular products, and the less popular products which are liked by only a few individuals will not be recommended. To fix this problem, we use a deep neural network that will find the hidden relations between items and users. These relations could be among different items or among different users or simply between different items and users. These hidden relations improve personalized recommendations and increase the overall accuracy of our model.

In our approach, the following is employed:

1. A User vectorization function, which creates a user data vector that incorporates the users similarity based on the user-user network and user-user clusters. Clusters are created using Expectation - maximization clustering. The output of this function is used as an input for the neural network engine(NNE).
2. An Item vectorization function, which uses collaborative filtering and a movies' similarity function, is employed to provide item vector. The output of this function is also used as an input for NNE.
3. The Neural Network Engine: it takes input from user and item vectorization functions and provides final ratings as the output. It also takes domain attributes as input to deal with the cold start problem.

We perform regressive experimentation using Flixster Dataset [7, 38], MovieLens 10M Dataset [13], and IMDBPy web crawler [17]. We achieved upto 27% improvement in the accuracy of predictions when compared against user and item-based CF algorithms. We also achieved up to 19.16% improvement in rating prediction in comparison to other Hybrid Recommender models. We also compare the running time of different popular hybrid RS approaches against the running time of our approach.

The paper is structured as follows: Section 2 contains a literature review on previously employed and related techniques. Section 3 explains our approach in detail with relevant diagrams. Section 4 is the experimentation section; it contains dataset and data preparation information as well as thorough experimental analysis. Section 5 concludes our research.

We share our experimental achievements and explanation in brief, along with future work and applications of our approach in Section 5.

2 Related work

This section contains a brief background on different RS approaches, deep learning based RS approaches and Movie RS.

2.1 Recommender system approaches

RS is classified according to the strategy used in the recommendation process. A lot of techniques have been proposed for RS,but three main techniques used are collaborative filtering techniques [34], content-based [32] techniques and Hybrid techniques [13].

2.1.1 Collaborative filtering

Collaborative Filtering (CF) method is based on the principle that two persons who have agreed in the past, may in the future agree again [12, 34]. Therefore, CF method first searches users with a taste similar to that of the target user. CF RS then gives suggestions to the target user by anticipating the target user's rating for specific target item based on the ratings of the user's top-S alike users. The ratings of users are generally represented by values within a specific range, e.g., 1 - 5. CF-based techniques are widely used for web-based services, such as Amazon, Facebook, Twitter, etc [12]. CF systems are impaired by cold start [19] problems. The Cold start problem arises in situations when the system cannot draw any inferences for users or items due to insufficient information.Other drawbacks of Collaborative Filtering recommender systems include sparsity in the rating matrix. Various dimensionality reduction techniques like SVD [4] & PCA [4] have been used to deal with sparsity.

Collaborative Filtering can be classified majorly into two categories: Memory-Based Approaches [37] and Model-Based Approaches [28].

Memory-Based Collaborative Filtering uses all the user-item data available. They establish their recommendations on the principle that similar users are more likely to have similar tastes, so their recommendations should also be similar. It is a community-goaded technique. Memory Based Recommendation according to Kai Yu et al. [37] can be further divided into Item-based technique and User-based technique. In User-Based technique, users who rated the items similarly may have same taste. Based on this , it recommends new items which are rated highly by similar users. Item Based technique focuses on items; it assumes that items which are rated similarly are presumably similar, and therefore, it recommends items with the highest similarity.

Model Based Collaborative Filtering [6, 28] gives item recommendation using a pre-constructed model. The model is built by various ML techniques such as Bayesian networks, clustering, etc. These models are also susceptible to cold-start and sparsity problems.

2.1.2 Content based approaches

The content-based techniques are developed for learning the potential interests of users based on the content and domain data of users or items [32]. Due to privacy reasons, collecting the information of users for RS is not easy. Hence, content based recommender

systems mostly use domain specific knowledge to create feature representations. For example - domain specific information for movies RS might use genre, year of release, actors, language, director, etc. All the feature representations are hand-engineered. Therefore, content-based approaches can give recommendations using only the existing preferences of the user. If users' preferences expand or change, the feature representations will need to be changed in the future. Jeevamol [19] used a hybrid ontology-based system to alleviate the cold start problem. Ontology is a detailed proper specification of a shared conceptualization. Jeevamol [19] used it in e-learning RS, where ontology was used to model the information about the learning resources and the learner.

2.1.3 Hybrid approaches

Hybrid approaches [13, 19, 23] combine two or more approaches to create a new hybrid approach. Most hybrid approaches [19] combine both Collaborative filtering and Content based approaches. Hybrid approaches usually gives a higher predictive accuracy than content based and collaborative filtering approaches. V kumar & A Tanwar [23] presented a hybrid folksonomy-based recommender system which combines domain knowledge of the items with the collaborative process. They used folksonomy to incorporate the domain data. Folksonomy is a tag based system which links various users and items with the help of tags. To improve the accuracy of the system they used fuzzy clusters to represent the neighborhood in a probabilistic manner. Dan yang [36] combined CF with a temporal association rule-based algorithm that can find users' hidden interests with time-specific leveraging historical behavioural data without using domain knowledge. The Hybrid temporal RS evolves features of users' historical feedback. Then time-aware users' alike neighbors data is selectively processed, and time-aware item ratings are predicted to keep trail of the dynamics of users' choices.

2.2 Deep learning approaches

Recently, because of their stalwart learning capabilities, deep learning approaches [10] have been successfully implemented in various areas of Natural Language Processing [25], Audio Recognition [8], and Computer Vision [14]. Also, many attempts have been made to implement these deep learning(DL) techniques in recommender systems. Restrictive Boltzmann Machines [24] and Autoencoders [9] have also been used for RS. The main idea of these techniques is to rebuild the ratings of users by learning hidden arrangements in the historical explicit ratings. Self learning networks [31] can also be used for RS, these networks improves the accuracy of recommendation by getting implicit feedback from previous recommendations. Xin Xia [35] built a self learning network based RS which used feedback from previous sessions, to improve the recommendations in next session.

The DNNRec [22] utilizes embeddings for describing items and users to learn non-linear latent factors. Their approach uses embeddings and deep neural networks to find latent relations between users and items. Second, it retains traditional collaborative filtering techniques by learning latent factors and uses them to learn non-linear latent factors as embeddings. In this approach while, creating latent relations the domain structure of the particular domain is not considered. As a result the accuracy of recommendations is affected.

2.3 Movie recommender systems

Movie Recommender systems [5, 10, 23] recommend top-n suitable movies from a collection of titles to a user. Movie RS are one of the most explored fields in RS because the domain information and data sets for movie RS are readily available. Forouzandeh [10] used group classification along with ensemble learning to increase movies prediction accuracy in recommender systems. The approach classified most similar target users using a mixture of ensemble learning, fuzzy rules, and decision trees. Chauhan [5] captured emotion to recommend the relevant movies. Chauhan used CNN to recognize each facial micro expression and recommend the movie from a specific genre. Their approach also used decision trees and boosting algorithms to improve the predictive accuracy of movies recommendations. This approach suffers from subjective generalization problem.

In the next section, we explain our deep learning based hybrid RS approach in detail.

3 Deep neural network-based hybrid recommender system with user-user networks

Our approach is a hybrid RS approach. It uses collaborative filtering along with content based approaches to handle the cold start and sparsity problems. Our approach uses User-user networks to provide optimum collaboration and synergy factor. Although, improving the collaboration and synergy factor may lead to poor personalized recommendations in case of unpopular items. To overcome this, our approach uses deep learning network to find the latent relations among the unpopular items and users. Our Neural Network-based Recommender system approach is divided into three phases User-user Networks clustering (UNCP) phase, Collaborative filtering phase (CFP) and Neural networks engine phase(NNEP). In the UNCP, we take the user data from the database and create user-user networks and user-user clusters. Then we use the (User Networks Cluster) UNC function to calculate user network cluster similarity. In the CFP, we use vectorization methods on the rating matrix and calculate user vector and item vector. Then we use the output of UNCP and CFP as the input of NNEP. After processing in the neural network engine phase, we get target ratings matrix as output. A basic architecture of our approach is demonstrated in Fig. 1.

3.1 User networks clustering phase

This UNC phase calculates the user-user networks and user clustering factors of our Neural network embeddings and is further divided into three parts: User-user Networks part, the User clustering part and a third UNC vector part. In User cluster part we perform Expectation-maximization clustering for making clusters. In the User-user networks part, we create an Undirected acyclic graph from the user-user adjacency matrix. We use User-user network along with clustering instead of just using clustering because, using user-user networks, we can find relations between users belonging to different clusters. If we only used clustering, then to find the relation between two users belonging to different clusters, we will have to use the relation between the two clusters the users belong to, affecting the similarity between those users. Rather than giving generalized recommendations like other methods, our technique could give more specific recommendations.

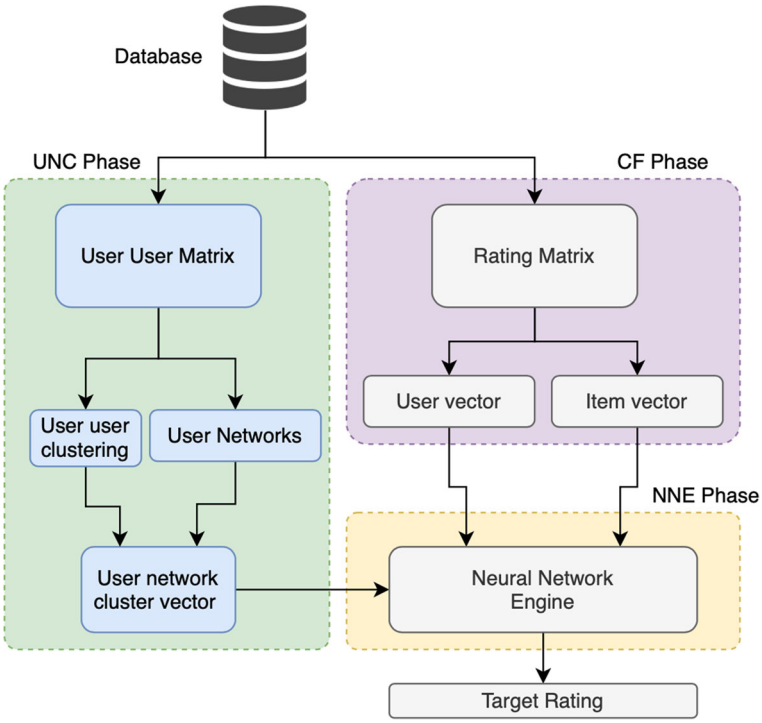


Fig. 1 Architecture of our deep neural network-based hybrid movies recommender system with user-user networks

3.2 User-user networks

In this part, we create an undirected acyclic graph with the User-user adjacency lists we have. The graph node of user 'i' in the graph is represented by G_i where

$$G_i => \begin{cases} user_id : i \\ neighbours_list : [n_{i1}, n_{i2}, n_{i3}, n_{i4}, ...] \end{cases} \quad (1)$$

here user_id represents the user id of the user *i*, neighbours_list represent the list of neighbours of *i*, n_{i1},n_{i2}, etc represent the user ids of the neighbouring nodes of *i*.

Each node G_i is unique. The weight of all edges is 1. Figure 2 represents a user-user network. We use a version of Loop free routing(LFR) [11] which uses Breadth First Search [27] to calculate the shortest path between all the nodes. Our modified LFR is faster than the traditional shortest path finding algorithms and can quickly update values when new nodes or edges are introduced. For calculating and updating the shortest paths, each node G_i is represented by g_i, where g_i includes two routing arrays: *dist* and *pred*, along with node G_i. g_i is represented as:

$$g_i => \begin{cases} node : G_i \\ dist : [G_{i1}, G_{i2}, G_{i3}, G_{i4}, ...] \\ pred : [G_{i1}, G_{i2}, G_{i3}, G_{i4}, ...] \end{cases} \quad (2)$$

dist[i1] stores the distance of node *i1* from the source *i* and the array *pred[0, ..., n-1]* such that *pred[i1]* represents the direct predecessor of the node *i1* in the BFS starting from *i*.

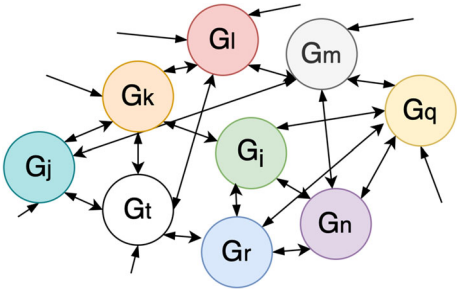


Fig. 2 User-user networks undirected graph

The shortest distance between two nodes is calculated using the Algorithm 1 and the BFS used in Algorithm 1 could be calculated by using Algorithm 2.

Require: $adj, src, dest, V$
  $pred_{list} \leftarrow empty\ array\ of\ size\ V$
  $dist_{list} \leftarrow empty\ array\ of\ size\ V$
  if $!BFS(adj, src, dest, V, pred_{list}, dist_{list})$ **then**
    $pred_{list}, dist_{list}$
  end if
  $path \leftarrow emptylist$
  $crawl \leftarrow dest$
  $path.add(crawl)$
  while $pred_{list}[crawl] \neq 1$ **do**
    $path.add(pred_{list}[crawl])$
    $crawl = pred_{list}[crawl]$
  end while
  return $pred_{list}, dist_{list}$

Algorithm 1 Modified LFR algorithm.

We calculate the similarity $\vartheta_{i,j}$ between two users U_i and U_j as

$$\vartheta_{i,j} = \frac{1}{s(U_i, U_j)^r} \tag{3}$$

Where s(U_i, U_j) is the shortest distance between users U_i and U_j in the user-user network graph. r is a parameter that determines the gradual decrease in the similarity value. $0 < r < 1$. When r = 1, the similarity is inversely proportional to the shortest distance.

3.3 User clustering

In this part, we use expectation-maximization(EM) [20] clustering over other clustering methods like K-means because EM clustering becomes generative and it accounts for variance as well. In EM clustering, we have two steps (1) the Expectation step and (2) the Maximization step. In expectation step The likelihood of user U_i w.r.t user U_j is given by Lik(θ, U_i, U_j) where

$$Lik(\theta, U_i, U_j) = p(U_i, U_j \mid \theta) \tag{4}$$

```
Require: adj, src, dest, V, pred, dist
  queue ← empty LinkedList
  visited ← empty array of boolean of size V
  for i = 0; i < V; i = i + 1 do
      visited[i] ← false
      dist[i] ← max integer value
      pred[i] ← −1
  end for
  visited[src] ← true
  dist[src] ← 0
  queue.add(src)
  while queue.size() ≠ 0 do
      u ← queue.dequeue()
      for i = 0; i < adj.get(u).get(i); i = i + 1 do
          if visited[adj.get(u).get(i)] == false then
              visited[adj.get(u).get(i)] ← true
              dist[adj.get(u).get(i)] ← dist[u] + 1
              pred[adj.get(u).get(i)] ← u
              queue.add(adj.get(u).get(i))
              if adj.get(u).get(i) == dest then
                  return true
              end if
          end if
      end for
  end while
  return false
```

Algorithm 2 BFS (Breadth First Search Algorithm).

$$p(U_i, U_j \mid \theta) = \int p(U_i, U_j \mid \theta) dU_j = \int p(U_j, U_i \mid \theta) dU_j \, p(U_i, U_j \mid \theta) dU_j$$

$$Lik(\theta, U_i, U_j) = \int p(U_j, U_i \mid \theta) dU_j \, p(U_i, U_j \mid \theta) dU_j \tag{5}$$

Where, θ is the unknown parameters vector. The Expectation step is denoted as

$$q^t = arg\ max_q Lik(\theta, U_i, U_j) \tag{6}$$

Where q is the arbitrary probability distribution and is chosen to maximize *Lik* The Maximization step is denoted as

$$\theta^{t+1} = arg\ max_\theta Lik(\theta, U_i, U_j) \tag{7}$$

Where θ is chosen to maximize *Lik* italic. The user cluster similarity $\varphi_{i,j}$ is then given by (8)

$$\varphi_{i,j} = \frac{1}{\left\{Lik(\theta, U_i, U_j)\right\}^r} \tag{8}$$

Where, r is a parameter that determines the gradual decrease in the similarity value. $0 <$ r < 1. When r = 1, the similarity is inversely proportional to the shortest distance.

3.4 User networks and clustering vector

In this part we just add and create a UNC vector for NNEP embeddings. The UNC vector λ_{UNC} is given by the (9)

$$\lambda_{unc} = \beta \times \varphi_{i,j} + (1 - \beta) \times \vartheta_{i,j} \tag{9}$$

Where β is the the weighting factor for combined user ratings.

3.5 Collaborative filtering

Collaborative Filtering is a method that requires only the data on historical preferences for a given assortment of items. As CF is based on historical preferences, it is based on the principle that two persons who have agreed in the past may in the future agree again [12, 34]. User preferences can be represented by two categories: Explicit Ratings and Implicit Ratings. Explicit ratings are the direct ratings given by the users on a specific scale; for example - A user may rate "The Avengers" 4 stars on a scale of 1-5 stars. Implicit ratings are indirect factors from which the user rating may be judged; for example - click-through rate, views, retention time, etc.

Rating matrix - rating matrix is a $n \times m$ matrix of ratings, where 'n' represents the number of users and 'm' represents the number of items. This matrix shows the rating of a user U_i for an item I_j, where $0 < i < n$ and $0 < j < m$. If a U_i has not rated I_j, then the cell (i,j) in the matrix will be zero or empty.

In this phase we use a technique of CF called Nearest Neighborhood Algorithm [2]. We use this technique to calculate both users similarity and items similarity. For calculating user similarity, We take an $n \times m$ rating matrix, with user $U_i (0 < i < n)$ and item $I_j (0 < j < m)$. Now to predict the rating $R_{i,j}$, we calculate the similarities among targeted user i and other users, then pick the top S similar users and use the weighted mean of ratings from these k users, including similarities as weights as shown in (10).

$$R_{i,j} = \frac{\sum_0^k Sim(U_i, U_k) R_{k,j}}{number\ of\ ratings} \tag{10}$$

Where Sim is the similarity, Similarly for Items

$$R_{i,j} = \frac{\sum_0^k Sim(I_i, I_k) R_{k,j}}{number\ of\ ratings} \tag{11}$$

While distinct users may have distinctive baselines when providing ratings, some users usually give higher scores, while some give lower scores even when they are pretty satisfied with the item. To overcome this behavior, we deduct each user's mean rating of all items when calculating the weighted mean and we add it back for the targeted user, as shown below.

$$R_{i,j} = \bar{R}_i + \frac{\sum_0^k Sim(U_i, U_k)(R_{k,j} - \bar{R}_k)}{number\ of\ ratings} \tag{12}$$

Similalrly for Items

$$R_{i,j} = \bar{R}_i + \frac{\sum_0^k Sim(I_i, I_k)(R_{k,j} - \bar{R}_k)}{number\ of\ ratings} \tag{13}$$

We use pearson correlation to calculate similarity as shown in (14).

$$Sim(U_i, U_k) = \frac{\sum_0^j (R_{i,j} - R_i)(R_{k,j} - R_k)}{\sqrt{\sum_0^j (R_{i,j} - R_i)^2 \sum_0^j (R_{k,j} - R_k)^2}} \tag{14}$$

Similarly for items,

$$Sim(I_i, I_k) = \frac{\sum_0^j (R_{i,j} - R_i)(R_{k,j} - R_k)}{\sqrt{\sum_0^j (R_{i,j} - R_i)^2 \sum_0^j (R_{k,j} - R_k)^2}} \tag{15}$$

But, there are some limitations of CF. It suffers from sparsity in data and Scalability, so before performing the above steps we use Singular value decomposition to decompose the initial sparse rating matrix to lower-dimensional matrices with latent factors. SVD works on the principle of ternary factorization of a real-valued matrix as shown in (16)

$$R_{m\times n} = P_{m\times m} \times K_{m\times n} \times Q_{n\times n} \tag{16}$$

Where P and Q are orthogonal matrices and K is a diagonal matrix with elements $a0 \geq a1 \geq a2...an$. Studies show that removing the small singular values "noise" certainly heightens the consumer-item relationship. In our method, we use the mean rating for every user i as r_u and replace every non-zero rating r_{ui} with $r_{ui} - r_u$. Next we use SVD on the rating matrix R to obtain the decomposed matrix R_1. The Factorized rating matrix PR is calculated using the (17):

$$PR_{ui} = r_u + R_{1_{ui}} \tag{17}$$

After using the decomposed rating matrix for calculating the user and item similarity, we use the similarity values along with Movie attributes data in the Neural Network Engine.

3.6 Neural network engine phase

To utilize information regularizing the creation of latent models for items and users in matrix factorization, we present a novel Neural network engine(NNE) to learn the distribution representation of items or users in a bottom-up fashion comparable to the approaches of Sainath et al., 2015 [29] & Tang et al., 2015a [33]. Figure 3 represents a basic flow diagram of the NNE.

We use tensorflow in our NNE phase. The Graphics processing unit used to train our different Neural networks was Nvidia GTX 1060. We first use Convolution Neural Network (CNN) to get the complex rating representations, then we use Deep Learning Neural network created by complex rating representations of all User item pair and CNN to get our final ratings for the Target Ratings matrix (Fig. 4).

3.6.1 Convolution neural network

User vector, Item vector, and user cluster network vector acts as embeddings for our Neural Network. We also take Movie (Domain) information in our Embeddings. Word Embeddings are created using Keras embedding layer. Each embedding in NNE is represented as (18):

$$E_{i,j} = \vec{\lambda}_{unc(j)}, \vec{Sim}(I_i, I_j), \vec{Sim}(U_i, U_j), \vec{e}_{j1}, \vec{e}_{j2}, ... \tag{18}$$

The vector $\vec{e}_j$ represents the vector of an attribute of a movie 'j'. The attributes used could be director, genre, year, language, etc. The vectors are used as the embeddings for NNE, and for each user i and item j it is denoted by '$E_{i,j}$'. Now we use those embeddings in the convolution layer as inputs.

In the convolution layer, each neuron uses a filter known as convolutional kernel. We use rectified linear unit (ReLU) activation function to circumvent the issue of vanishing gradient [1]. Convolution product is represented as (19)

$$g_N = \delta(E \times G_N) \tag{19}$$

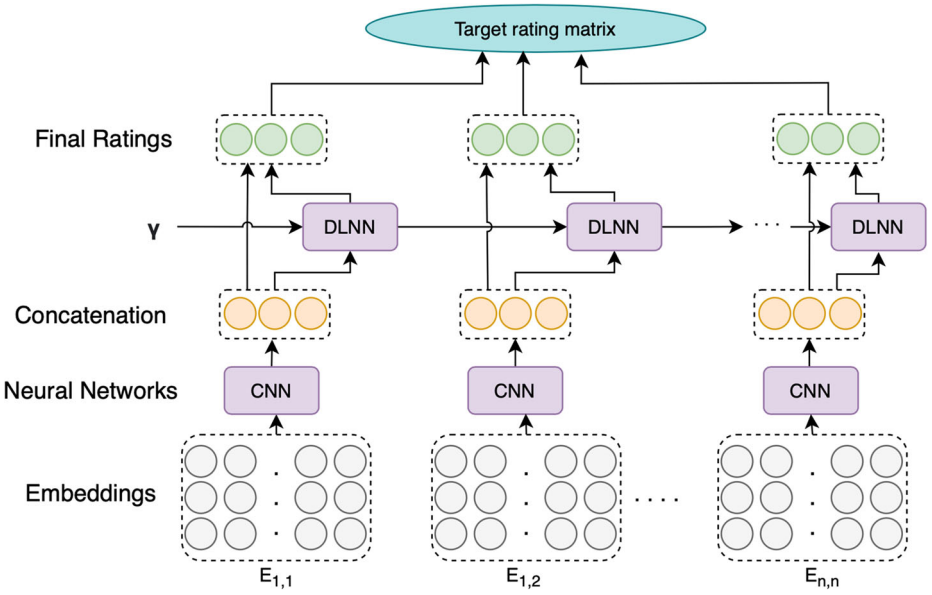


Fig. 3 Graphical representation of neural network engine

Where δ is an activation function and g_N is the convolution product for neuron N. After convolution layer's operations, every filter will build a feature map. For a specific kernel g_N, we next use a max-pooling operation on the above feature map and use the maximum value as the feature. The most notable feature of every feature map, which has the maximum value, is stored in a univariate vector o_N represented as (20)

$$o_N = max(g_1, g_2, ...) \tag{20}$$

Since one feature is extracted from each kernel, the numerous features derived from various kernels are concatenated to form a single feature C_t where

$$C_t = (o_1, o_2, o_3...) \tag{21}$$

We moreover use an activation method, hyperbolic tangent (tanh), to include pointwise nonlinearity [33] . We then concatenate all feature vectors generated to make complex ratings representation(CRR). For example if items $\in$ 1, 2, 3,.., we derive complex ratings representations by (22)

$$CRR = [tanh(C_1), tanh(C_2), tanh(C_3), ...] \tag{22}$$

3.6.2 Deep learning neural network (DLNN)

In DLNN phase, neural networks are created by using all complex rating representations created in the previous step. The DLNN first learns a representation for each CRR in the documents of users/items with convolutional neural networks (CNNs). Afterward, the interpretation of all CRR and their relations are adaptively encoded in document representation.

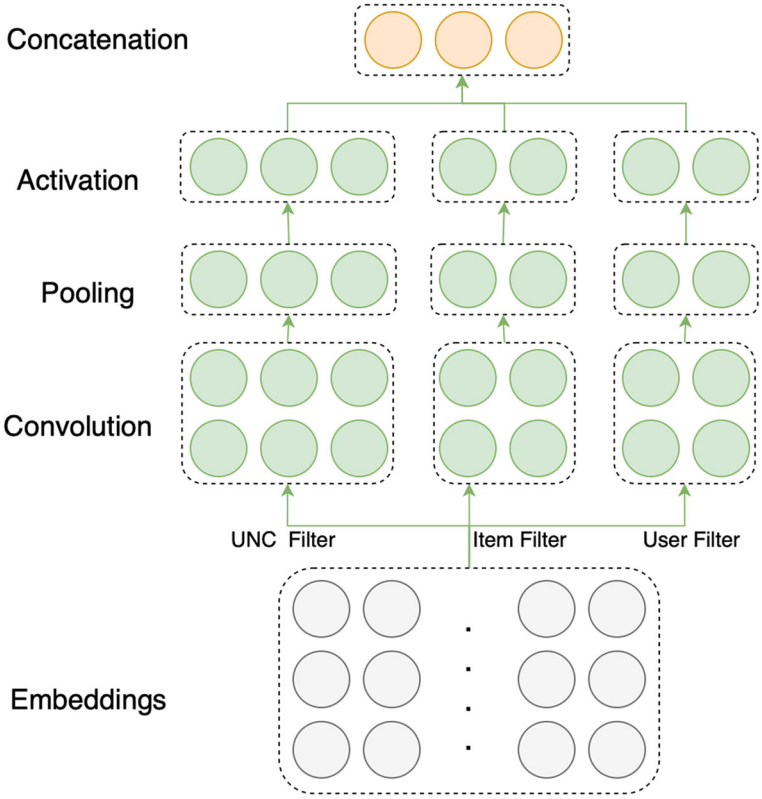


Fig. 4 Convolution neural network generating CRR

Finally, to calculate the final ratings, the complex rating representations and Deep learning ratings are combined to form the final ratings (F) given by (23).

$$F_{i,j} = \frac{CRR_{i,j} + \gamma \times (DLR_{i,j})}{2} \tag{23}$$

Where DLR is the deep learning representation and gamma is the Rating Matrix constant. The rating matrix constant is calculated by (24)

$$\gamma = \frac{log(n^{R}_{attr})}{n_{attr}} \tag{24}$$

Where n_{attr} represents the number of domain attributes considered in NNE phase and R represents the highest rating possible. In our case R would be 5.

Finally the final ratings are put into the Target rating matrix and are normalized to match the scale of the ratings.

In the next section we compare our approach with other RS approaches and analyze the findings.

4 Experimentation & analysis

For the evaluation of our model, we have considered the Flixster dataset [7, 38] & MovieLens 10M dataset [13].

Flixster is a social movie site that allows users to share movie ratings, find new movies and engage with others having similar movie preferences. This dataset contains the friendship network crawled in December of 2010 by Javier Parra. The Flixster dataset was cleaned before use, according to our requirements. The Flixster dataset is a commonly used dataset in the recommendation domain consisting of 2,523,386 users and 9,197,338 connections among the users according to their shared preferences. The movie ratings given by these users are on a scale of 0-5 and are stored in a user-item rating matrix. MovieLens 10M dataset has about 10 million movie ratings and 95600 tags used on 10600 movies rated by 71500 users. We have also used IMDBpy [17], a python based web crawler for extracting users and movies data from IMDB. We randomly split the data into 80% for training and 20% for testing purposes. We chose two datasets to ascertain two things: firstly, our model is not overfitting a single dataset, and secondly, our model is adaptable to different datasets.

4.1 Evaluating our model

The proposed Deep Learning-Based Recommender system using User-user networks model is evaluated and compared with the previously used Collaborative Filtering based, Hybrid, and Neural-Network based models. We have used accuracy metrics like Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) & decision support metrics like F1 for evaluation and analysis.

Accuracy metrics compare the recommended ratings with previously known ratings of user-item pair. Accuracy measures like MAE and RMSE between the estimated and actual ratings can be calculated by (25), (26):

$$MAE = \frac{\sum_{i=1}^{N} |PR_{u,i} - AR_{u,i}|}{N} \tag{25}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (|PR_{u,i} - AR_{u,i}|)^2}{N}} \tag{26}$$

where PR is the predicted rating, AR is the actual rating and N is the number of items rated by user u respectively.

Equation 9 uses the weighting factor to calculate the final similarity of the UNC phase. The lower the weighting factor, the higher the influence of user-user networks on UNC phase similarity and vice versa. By looking at the graph in Fig. 5, it can be seen that by decreasing the weighting factor in our model, MAE also decreases for both MovieLens and Flixster datasets. MAE decreases with a decrease in weighting factor till the weighting factor value of 0.3. After 0.3, MAE starts to increase. This increase in MAE result from the highly reduced influence of cluster similarity in the final UNC similarity. Therefore, to get the right balance between user-user networks and user clustering similarity, we take the weighting factor value to be 0.3. The conclusions from the graph in Fig. 5 show us that User-user networks combined with clustering provide a better collaboration and synergy factor than simply using clustering.

Considering the datasets used are quite large, the number of relevant items may appear negligible in some cases, and thus MAE and RMSE may appear misleading sometimes. So

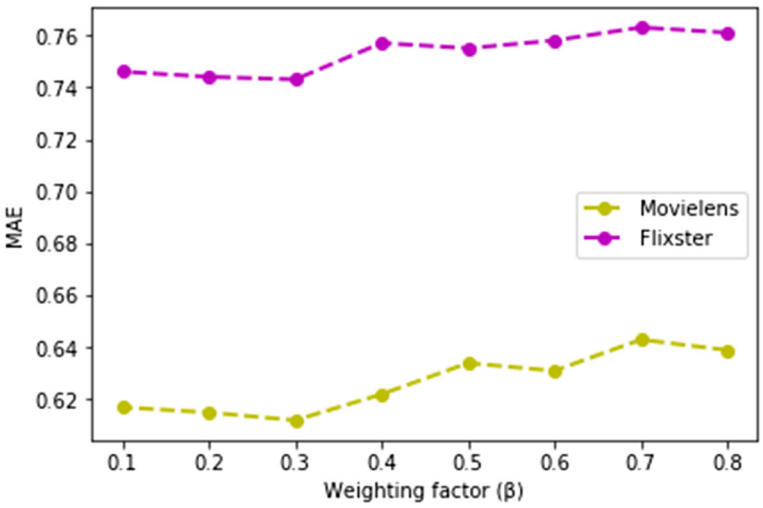


Fig. 5 Variation of MAE with change in Weighting factor

in order to resolve such imbalances in classification problems, we use decision support metrics such as Precision and Recall. The precision evaluates the relevance of the recommended items. On the other side, the recall evaluates the tendency to recommend all the relevant items. Precision and Recall are calculated using (27), (28).

$$Recall = \frac{|Rec \wedge Rel|}{|Rel|} \tag{27}$$

$$Precision = \frac{|Rec \wedge Rel|}{|Rec|} \tag{28}$$

where Rec and Rel are the sets of recommended and relevant items respectively.

A balance between recall and precision is obtained by computing the F1 score as the harmonic mean of precision and recall.Very small precision value or recall value results in a lower overall score. Thus it aids in balancing the two metrics. F1 is given by (29):

$$F1 = \frac{2.Recall.Precision}{Precision + Recall} \tag{29}$$

We computed the RMSE, MAE & F1 score of different Recommender System approaches on the MovieLens 10M and Flixster Dataset. Table 1 shows our experimental findings on MovieLens 10M Dataset & Table 2 shows our experimental findings on Flixster Dataset (Fig. 6).

We first calculated F1 score for User based Collaborative filtering (UCF) and Item based Collaborative Filtering (ICF) Models. Then we applied SVD on both UCF & ICF models. We then calculated the metrics for Hybrid Recommender System model which uses both user CF and item CF along with SVD [26]. Lastly, We compared our model with another state of the art recommender system called Deep Learning based Collaborative Neural Networks (DLCNN) RS which also uses Neural Networks [3].

User based Collaborative Filtering (UCF) model is basic collaborative filtering model, which find similarity between users using cosine similarity. UCF then uses this similarity to recommend top-n items. Similarly, Item based collaborative filtering model is also a basic CF model which finds similarity between items using cosine similarity. ICF uses the

Table 1 Evaluation and comparison of Our Model with various approaches on MovieLens 10M Dataset

MovieLens 10M dataset [13]			
Recommender system approach	RMSE	MAE	Mean F1@50
User based CF(UCF)	0.936	0.713	0.610
Item based CF(ICF)	0.932	0.706	0.611
UCF with SVD	0.893	0.675	0.631
ICF with SVD	0.891	0.674	0.633
Hybrid user and item based with SVD [26]	0.886	0.666	0.638
DRL RS [15]	0.877	0.650	0.643
DLCNN [3]	0.842	0.651	0.641
Our approach	0.767	0.612	0.662

similarity calculated to recommend top-n items from top-m similar users. The early adopted recommender systems UCF and ICF have pretty high RMSE and MAE & very low Mean F1 score. This is because of the fact that CF models suffer from sparsity and cold start problems. To tackle the sparsity problem, we apply dimensionality reduction techniques. So after applying single value decomposition, the complexity of the matrix was reduced and by using latent factors for ratings prediction, the accuracy as well as classification improved for UCF model with SVD and ICF model with SVD as seen in Fig. 7. But the accuracy and prediction values for these models are still low. This is because these models still suffer from cold start problem. Therefore, To improve the accuracy and solve the cold start problem, Nilashi et al. [26] combined UCF with SVD model, ICF with SVD model and content based approaches to create a hybrid user and item based with SVD recommender system. The Hybrid RS [26] The accuracy and classification metrics improved considerably compared to using just CF models.

The Hybrid RS model used Ontology [19] in the content based system. Ontology is a formulation of a particular field like movies, courses, construction materials, etc into a human-apprehensible, but computer understandable form consisting of entities and attributes along with their various relationships and axioms. Ontology helped Hybrid RS ameliorate the cold start problem. The Hybrid CF RS model provides us better accuracy than CF models but the accuracy is still poor. This is because CF RS are biased towards popular items. Hence, Deep learning techniques [3, 15] are employed to improve the accuracy.

Table 2 Evaluation and comparison of Our Model with various approaches on Flixster Dataset

Flixster dataset [7, 38]			
Recommender system approach	RMSE	MAE	Mean F1@50
User based CF(UCF)	0.971	0.865	0.620
Item based CF(ICF)	0.980	0.860	0.618
UCF with SVD	0.921	0.853	0.633
ICF with SVD	0.915	0.857	0.629
Hybrid user and item based with SVD [26]	0.893	0.822	0.640
DRL RS [15]	0.896	0.812	0.642
DLCNN [3]	0.882	0.801	0.647
Our Approach	0.713	0.743	0.667

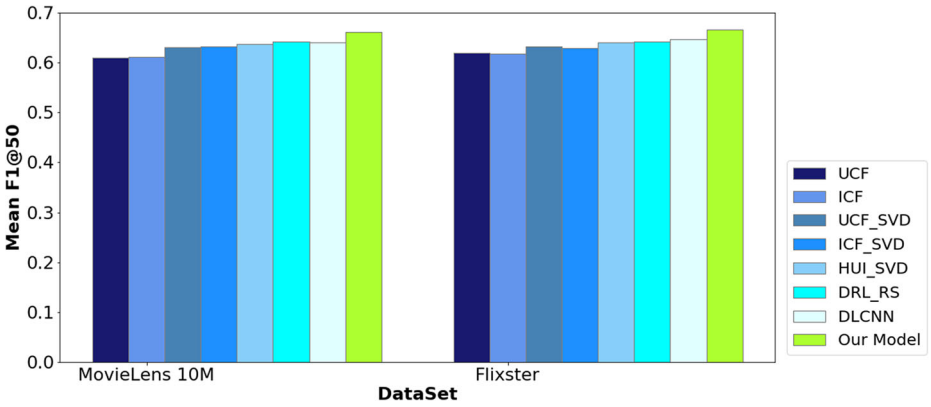


Fig. 6 Variation of F1 score compared to different RS approaches

Deep learning techniques find the latent relations among various items and users to improve the recommendations.

DLCNN [3] captures the deep and hidden interactions between users and items, by using two different feed forward neural networks along with Adam optimization. As seen in Tables 1 & 2, The DLCNN performed better than the Hybrid CF model, because the multistage deep collaborative neural network used in it models and discovers deep interactions which traditional CF based RS could not. The deep learning RS result in better personalized recommendation due to their ability to find these latent and deep interactions. Therefore, due to better personalized recommendations DLCNN is more accurate than Hybrid CF RS [3].

Most of the RS approaches including the aforementioned approaches are static in nature. User preferences evolve and change with time. Most recommender systems do not update with time, they need to be manually updated. A Deep reinforcement learning based long-term recommender system (DRL RS) [15] provides the solution to this problem by providing a self adapting reinforcement learning based RS. DRL RS uses recurrent neural networks and Markov decision processes to stimulate the reinforcement model based on the interactions between a user and the system. The DRL RS is a user oriented system, it does not take into account the domain structure and domain information. Due to this limitation, DRL RS provides poor domain-specific recommendation and the accuracy of DRL RS is less than our model and DLCNN.

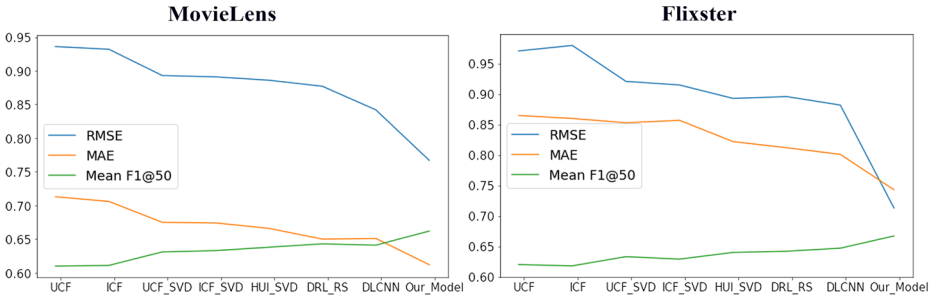


Fig. 7 Changes in accuracy and classification metrics of different RS approaches

Table 3 Run time of various RS approaches on different datasets

Model	DataSet	Mean (min)	S.Deviation (min)
Hybrid CF with SVD [26]	MovieLens 10M	132.22	11.56
DRL RS [15]	MovieLens 10M	621.21	22.57
DLCNN [3]	MovieLens 10M	474.55	20.37
Our approach	MovieLens 10M	397.43	9.84
Hybrid CF with SVD [26]	Flixster	92.41	2.35
DRL RS [15]	Flixster	491.47	17.61
DLCNN [3]	Flixster	350.89	14.24
Our approach	Flixster	303.56	6.35

Our models uses domain attributes in the NNE phase to overcome the cold start problem. Therefore our model gives better accuracy than basic CF models. Our model uses user-user network which provides better collaboration and synergy factor than that of DLCNN. But improving the synergy and collaboration factor in some cases may lead to poorer personalized recommendations, so we use deep learning to deal with the problems of improved synergy and collaboration factor. Hence, our model is able to find the right balance among personalized, domain-based and collaborative preferences. It is due to this balance that our model is able to outperform DLCNN, DRL RS and other state-of-the-art recommender systems.

We also compared the running time of different approach with our approach. Running time in our approach is measured by the standard deviation and mean across nine runs.

Table 3 presents the mean run time over nine runs (in minutes) of various approaches on MovieLens 10M and Flixster Dataset. The run time also includes the graph creation time, cluster creation time & Neural network generation time. DLCNN, DRL RS and our approach are not faster than Simple Hybrid CF with SVD approach, as DLCNN and our approach are more complex, meticulous and accurate. DRL RS takes the longest time, because it is a reinforcement learning based technique and it takes longer to update the whole deep learning network after each interaction. DLCNN takes longer time than our approach because it performs collaborative filtering in multiple stages and feature representations using matrix factorization separately. The Table 3 proves that our model is faster than other state-of-the-art neural network approaches while having better accuracy and classification metrics.

5 Conclusions & future work

In this approach, a neural network-based hybrid recommender system using user-user networks was introduced. Expectation maximization clustering was used to make user clusters and Pearson correlation was used to calculate collaborative filtering similarity. The user-user networks used provide a superior collaboration and synergy factor to our model. This was proved by comparing its accuracy with accuracy of other collaboration based models. But a great collaboration and synergy factor may lead to poorer personalized recommendations in some cases. To solve this the proposed model used a complex deep-learning neural network to provide a right balance among personalized, domain-based and collaborative approaches. It utilizes embeddings for learning non-linear latent factors of items and users.

It used domain embeddings to overcome cold-start problem. It unites the deep learning characteristics with user connections information available from user-user network to create a hybrid system. The introduced architecture combines content-based and collaborative filtering-based approaches to overcome the cold-start and overfitting problems.

The differences between our model and other deep learning based recommender systems are: (1) Our model uses user-user networks which provides it superior collaboration and synergy factor (2) Unlike other models which use the output from content based approaches in creating embeddings, our model uses the input of content based approaches in the neural network phase. Using the input afterwards doesn't distort the domain structure of domain attributes.

The implementation results of our method was compared with other RS techniques, which used collaborative filtering, clustering, Ontology, SVD, Neural networks, matrix factorization and reinforcement learning. All these recommender system approaches were analyzed on the MovieLens 10M and Flixster datasets. The approaches were analyzed on both running times and predictive accuracy. Our implementation performed better than other CF models, hybrid models and deep learning based methods in predictive accuracy. Our approach outperformed the previous approaches' RMSE by 19%, MAE by 9.2% & Mean F1 by 4.1%.

Our model could be used in multiple domains like E-learning where it could be used to recommend the right courses to students. The optimum use of our model will be in social network domain where users are connected to each other and can buy items right from the social media platform. But, due to unavailability of relevant data we cannot test the efficacy of our model on major social media platforms. Despite being faster than other Neural Network-based RS, the running time of our model is higher than many RS approaches, including the running time of the Hybrid CF RS approach. This longer running time could be attributed to the complexity of our method, as we calculate numerous attributes during each phase. Moreover, due to the nature of undirected graphs, the addition of each new user extends the processing time. The Neural networks also take a considerably long learning time, even with a dedicated GPU. In the future, we will address these limitations by introducing simpler hash tables for user-user networks and better Neural network embeddings and gated architecture.

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

Declarations

Financial Interests No competing financial interests in this research.

Competing interests

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

References

1. Agarap AF (2018) Deep learning using rectified linear units (relu). arXiv:1803.08375
2. Ali N, Neagu D, Trundle P (2019) Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets. SN Appl Sci 1.12:1–15
3. Almaghrabi M, Chetty G (2018) A deep learning based collaborative neural network framework for recommender system. In: International conference on machine learning and data engineering (iCMLDE), IEEE, 2018

4. Anowar F, Sadaoui S, Selim B (2021) Conceptual and empirical comparison of dimensionality reduction algorithms (pca, kpca, lda, mds, svd, lle, isomap, le, ica, t-sne). Comput Sci Rev 40:100378
5. Chauhan S, Mangrola R, Viji D (2021) Analysis of Intelligent movie recommender system from facial expression. In: 2021 5th international conference on computing methodologies and communication (ICCMC), IEEE
6. Di Noia T, Mirizzi R, Ostuni VC, Romito D. (2012) Exploiting the web of data in model-based recommender systems. In: Proceedings of the sixth ACM conference on recommender systems, pp 253–256
7. Dong T (2017) Flixster-dataset.zip. figshare. Dataset. https://doi.org/10.6084/m9.figshare.5677741.v1
8. Ertam F (2019) An effective gender recognition approach using voice data via deeper LSTM networks. Appl Acoust 156:351–358
9. Ferreira D et al (2020) Recommendation system using autoencoders. Appl Sci 10.16:5510
10. Forouzandeh S, Berahmand K, Rostami M (2021) Presentation of a recommender system with ensemble learning and graph embedding: a case on MovieLens. Multimed Tools Appl 80.5:7805–7832
11. Gianlorenzo AD, Emidio MD, Frigioni D (2014) A loop-free shortest-path routing algorithm for dynamic networks. Theor Comput Sci 516:1–19
12. Hameed MA, Jadaan OA, Ramachandram S (2012) Collaborative filtering based recommendation system: A survey. Int J Comput Sci Eng 4.5:859
13. Harper F, Konstan JA (2015) Maxwell The movielens datasets: History and context. Acm Transactions on Interactive Intelligent Systems (tiis) 5.4:1–19
14. Hossain M, Muhammad G (2019) Shamim Emotion recognition using deep learning approach from audio–visual emotional big data. Inform Fusion 49:69–78
15. Huang L et al (2021) A deep reinforcement learning based long-term recommender system. Knowl-Based Syst 213:106706
16. Hwang T-G et al (2016) An algorithm for movie classification and recommendation using genre correlation. Multimed Tools Appl 75.20:12843–12858
17. Malagoli A, IMDBpy web crawler (2019), GitHub repository, https://imdbpy.github.io Accessed 12 August 2021.
18. Jazi SY, Kaedi M, Fatemi A (2021) An emotion-aware music recommender system: bridging the user's interaction and music recommendation. Multimed Tools Appl 80.9:13559–13574
19. Jeevamol J, Renumol VG (2021) An ontology-based hybrid e-learning content recommender system for alleviating the cold-start problem. Educ Inf Technol 26.4:4993–5022
20. Jung YG, Kang MS, Heo J (2014) Clustering performance comparison using K-means and expectation maximization algorithms. Biotechnol Biotechnol Equip 28.sup1:S44–S48
21. Katarya R, Arora Y (2020) Capsmf: A novel product recommender system using deep learning based text analysis model. Multimed Tools Appl 79.47:35927–35948
22. Kiran R, Kumar P, Bhasker B (2020) DNNRec: A novel deep learning based hybrid recommender system. Expert Syst Appl 144:113054
23. Kumar V et al (2021) Folksonomy based fuzzy filtering recommender system. In: 2021 IEEE symposium series on computational intelligence (SSCI), IEEE
24. Kuo RJ, Chen JT (2020) An application of differential evolution algorithm-based restricted Boltzmann machine to recommendation systems. J Internet Technol 21.3:701–712
25. Nguyen LV, Nguyen T-H, Jung JJ (2020) Content-based collaborative filtering using word embedding: A case study on movie recommendation. In: Proceedings of the international conference on research in adaptive and convergent systems
26. Nilashi M, Ibrahim O, Bagherifard K (2018) A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques. Expert Syst Appl 92:507–520
27. Rahim R et al (2018) Breadth first search approach for shortest path solution in Cartesian area. Journal of Physics: Conference Series. Vol 1019. No. 1 IOP Publishing
28. Sadaei HJ, Enayatifar R, Lee MH, Mahmud M (2016) A hybrid model based on differential fuzzy logic relationships and imperialist competitive algorithm for stock market forecasting. Appl Soft Comput 40:132–49
29. Sainath TN et al (2015) Convolutional, long short-term memory, fully connected deep neural networks. In: 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP) IEEE
30. Shaikh S, Rathi S, Janrao P (2017) Recommendation system in E-commerce websites A graph based approached. In: 2017 IEEE 7th international advance computing conference (IACC), IEEE
31. Srinivasu PN, Balas VE (2021) Self-learning Network-based segmentation for real-time brain MR images through HARIS. PeerJ Comput Sci 7:e654

32. Sulthana RA, Ramasamy S (2019) Ontology and context based recommendation system using neuro-fuzzy classification. Comput Electr Eng 74:498–510
33. Tang D, Qin B, Liu T (2015) Document modeling with gated recurrent neural network for sentiment classification. In: Proceedings of the 2015 conference on empirical methods in natural language processing
34. Thakker U, Patel R, Shah M (2021) A comprehensive analysis on movie recommendation system employing collaborative filtering. Multimed Tools Appl, 1–26
35. Xia X et al (2021) Self-supervised hypergraph convolutional networks for session-based recommendation. In: Proceedings of the AAAI conference on artificial intelligence, vol 35, no 5
36. Yang D, Nie ZT, Yang F (2021) Time-Aware CF And temporal association Rule-Based personalized hybrid recommender system. J Organ End User Comput (JOEUC) 33.3:19–34
37. Yu K, Schwaighofer A, Tresp V, Xu X, Kriegel H. P. (2004) Probabilistic memory-based collaborative filtering. IEEE Trans Knowl Data Eng 16(1):56–69
38. Zafarani R, Liu H (2009) Social Computing Data Repository at ASU [http://socialcomputing.asu.edu]. Tempe AZ: Arizona State University, School of Computing, Informatics and Decision Systems Engineering. Accessed 11 August 2021
39. Zihayat M et al (2019) A utility-based news recommendation system. Decis Support Syst 117:14–27

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

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

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3214348

A Graph Convolution Network based Adaptive Cooperative Spectrum Sensing in Cognitive Radio Network

Dimpal Janu, *Member, IEEE,* Sandeep Kumar, *Senior Member, IEEE,* and Kuldeep Singh, *Senior Member, IEEE*

***Abstract*—The hidden node problem is one of the most challenging issue in Cooperative Spectrum Sensing (CSS). The system models adopted by the existing Deep Learning-based spectrum sensing methods have not focused on modeling the hidden node scenario in cognitive radio networks. Further, these methods are unable to adapt to the dynamic channel conditions in the wireless environment since they have not considered the effect of fading environment. Motivated from these limitations, we propose GCN-CSS, a novel Graph Convolution Network (GCN) based cooperative spectrum sensing methodology which adapts to the dynamic changes in the Cognitive Radio Network. To the best of the author's knowledge, this is the first work to apply GCN for solving CSS problem. We have considered a practical system model which handles the dynamic channel condition i.e. SUs with multiple antennas experiencing different fading models with different fading severity. We have also catered the scenario of imperfect reporting channel between the SUs and the fusion centre along with the imperfect sensing channel to prove the robustness of the proposed model. With sufficient simulations, the superiority of the proposed methodology is proven in different dynamic scenarios of the wireless environment.**

***Index Terms*—Spectrum Sensing, Graph Convolution Network, Multi-antenna Cooperative Spectrum Sensing, Fading Channels.**

I. Introduction

DUE to the exponential proliferation in the usage of wireless traffic networks through the internet of things (IoT) and device-to-device (D2D) communication applications, a severe shortfall of the radio spectrum is being faced worldwide. The main reason for this underlying problem is the fact that the available spectrum resources are not being used consummately. Cognitive radio (CR) network is a promising solution to this subsisting predicament [1]. There are two types of users in the CR network, the licensed user which is the primary user (PU), and the un-licensed user which is the secondary user (SU) . SU can access the licensed spectrum when it is not being used by the corresponding PU, thus effectively abating the spectrum shortage issue. In the working of the CR network, SU has to be prudent in ensuring that zero interference is caused to the PU while accessing the spectrum of the PU. The biggest challenge in the network is to sense the presence of available spectrum and hence spectrum sensing (SS) is the utmost influential function in the CR network. There are various avenues used for SS and energy detection is the most widely studied in literature due to its simple circuitry and low complexity [2]. Due to the random nature of fading channels between the PU and the SU, the detection of the spectrum for a single node SU is susceptible to error. To overcome this challenge,CSS, where many SUs do the individual SS and contribute in making final decision has been used to improve the SS performance of the system.

However, designing the generic CSS strategy for different channel conditions is complicated as the scenarios for the SUs which are close to PU are different from those, which are far away from PU. The most popular CSS strategies used in the literature are ORing, ANDing, and K-out-of-N, etc., where the individual sensing of SUs is considered to be independent of each other. Although the traditional CSS methods show performance improvement over the individual sensing scheme, their performance degradation is entailed in the dynamic network environment.

Recently, various deep/machine learning (DL/ML) algorithms have been utilized to improve the performance of CSS. An extensive survey of the ML methods used for CSS is provided in [3]. The paper discusses the benefits as well as drawbacks of these ML based methods in terms of performance gains, complexity and applicable scenarios along with the details of the features extracted from the sensing signal. Few seminal works for solving spectrum sensing problems using ML are reviewed first and then the DL based solutions are discussed below. Awe *et. al.* in [4] applied error correcting output code based multi-class SVM algorithms to solve the SS problem, considered as multi state signal detection problem under multiple PU scenarios. To improve the sensing performance further, an information pre-processing scheme such as principal component analysis (PCA) has been utilized for main information extraction so that impact of interfering information on the detection results could be minimized, whereas genetic algorithm clustering is used for classification of spectrum states [5]. Xu *et. al.* [6] have designed a Bayesian ML-based framework to find global spectrum states, which exploits the mobility of SUs in order to enhance the spatial-temporal diversity of the received signal. The major drawback of the ML-based techniques is that they use manually extracted features as test statistics for decision making and these handcrafted features are not able to extract the hidden representations of the sensing data as well as can not accurately characterize the real environment. The whole ML model needs to be retrained every time whenever the

D. Janu and K. Singh are with the Department of Electronics and Communication Engineering, Malaviya National Institute of Technology Jaipur, 302017,India (e-mail: 2019rec9146@mnit.ac.in;kuldeep.ece@mnit.ac.in)

S. Kumar is with Central Research Lab, Bharat Electronics Ltd., Ghaziabad, 201010, India (e-mail: sann.kaushik@gmail.com)

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:56:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3214348

input signal feature undergo changes which is computationally complex and time inefficient process.

Convolutional Neural Networks (CNN) have shown great success in various fields since they have powerful capability to automatically learn various discriminative features from matrix shaped data. The CNNs have gained the interest of researchers for solving the spectrum sensing problem. The covariance matrix (CM) is considered as the versatile test statistic for representing the discriminative features obtained from sensing data, it is the popular choice for the input feed to the CNN. Liu *et. al.* [7] has proposed CNN-based test statistics for spectrum sensing considering CM as input. This work was extended in [8] where a cost function was derived as per maximum a posteriori probability (MAP) criterion and the CNN-based likelihood ratio-test was considered for SS according to the Neyman-Pearson rule. Chen *et. al.* [9] considered the time-frequency matrix of sensing signal obtained from short Time Fourier transform (STFT) analysis as the input to the CNN-based model. Zheng *et. al.* [10] designed a CNN-based non-cooperative SS framework to enable the high-quality spectrum predictions at different locations, frequency bands, and SNR environments by making use of transfer learning. The above discussed methods have utilized CNN architecture for extracting spatial features of the sensing signals, however, researchers have also focused on extracting temporal features by exploiting the past sensing data. Hence, to further improve detection accuracy, authors of [11] have used a CNN-based algorithm and exploited historical sensing data along with present sensing data to learn the PU activity pattern.

Few related works in literature have utilized Long Short Term Memory (LSTM) network, which is capable to extract temporal features by exploiting past sensing data of multiple sensing periods. A SS framework based on LSTM network has been developed [12], in which past sensing data is fed along with the present time sensing data to LSTM cell to exploit temporal dependencies in the data. Soni et. al. in [13] have proposed a SS model based on LSTM network which learns the implicit features from sequential spectrum sensing data i.e. time correlation between the present sensing event and historical sensing event. Along with this, the statistical information of PU activity such as ON and OFF period duration of PU, and duty cycle are utilized to improve the sensing performance of the CR network. In [14] Taguchi method is employed to determine the best optimized configuration of the neural network(NN) and analyze the effect of hyperparameter on LSTM network for spectrum prediction.

The combination of CNN and LSTM networks has the powerful capability to extract spatial and temporal features from the sensing data which can further enhance the spectrum sensing performance. In [15], authors have developed a SS framework based on CNN and LSTM models, where CNN is utilized to extract energy correlation features from the sensing data, and then these series of energy correlation features corresponding to multiple sensing periods are fed to the LSTM network to extract temporal features for learning the activity pattern of PU. In [16], blind spectrum sensing model was proposed based on one dimensional CNN and LSTM, to handle the low signal-to-noise ratio scenario. Ke et. al. [17] experimented with the combination of CNN and LSTM to extract the frequency and time-domain features of the sensing data for PU detection in a low SNR environment.

The DL-based methods proposed earlier have not considered the effect of different fading environments in the CR network. Thus, they are unable to adapt to the dynamic channel conditions in the wireless environment. Further, the hidden node problem which is a prominent hindrance in cooperative spectrum sensing is also not addressed effectively by these approaches. The DL models such as CNN is effective in extracting latent representations from the Euclidean data such as image, text, and videos. The data samples to be fed to CNN are represented as a regular grid in Euclidean space. The existing literature on CNN-based CSS assumes that the sensing signal i.e. input to the CNN can be represented in Euclidean space as fixed-size matrices. However, in the case of variable number of hidden nodes, these assumptions cannot be true always and the variable size of sensing matrices has to be catered for decision making at the fusion centre. In this scenario, it is evident to represent the variable size sensing matrices in non-Euclidean space. Hence the sensing matrices can be converted into graphs with complex relationships and interdependency between objects.

The ML & DL based CSS methods discussed above have assumed that the reporting channel between the SUs and the fusion centre is perfect wireless channel. However, in practical scenarios this channel is prone to different fading conditions [18]. The existing literature is failed to address the challenges associated with the imperfect reporting channel condition.

To overcome these limitations, we propose a novel CSS methodology that is robust to the dynamic changes in the wireless environment of the CR network. Recently, GCN [19] have gained increasing attention, since it can encode the structural information of underlying non-Euclidean or non-grid like data to model the relationship among entities. GCN has shown promising results in the field of remote sensing for hyperspectral image classification as the data is available in non-grid structure [20]. Inspired by the powerful ability of GCN, we employ it to model the relation between the primary user signal sensed by different secondary users scattered over a geographical area using graph data in non-Euclidean space. The variation in the number of hidden nodes is modeled through graphs and the GCN is capable of processing these variable size graphs.

The novel contributions of this paper can be summarized as follows.

- We propose a novel GCN-based CSS (GCN-CSS) scheme, capable of adapting to the dynamic changes in the wireless environment. We utilize GCN to model the potential relations between signals sensed by the secondary users, providing more discriminative information for deducing the presence or absence of the primary user signal. Although GCN has been extensively utilized in different data settings, to the best of the author's knowledge this is the first work to apply GCN for solving CSS problems.
- We have considered a practical system model which handles the dynamic channel condition i.e. SUs with

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:56:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3214348

multiple antennas experience different fading models with different fading severity. We have also catered the scenario of imperfect reporting channel between SUs and the fusion centre along with imperfect sensing channel to prove the robustness of the proposed model. As per the best of our knowledge, this generalized dynamic practical scenario is being considered for the first time to solve the CSS problem.
- The proposed GCN-CSS method addresses the hidden node problem which is one of the most challenging issues in the spectrum sensing.
- We have validated that the proposed scheme outperforms all the conventional and recently proposed schemes in the dynamic wireless environment scenario considered here.

The remainder of this article is organized as follows. The scenario of multi-antenna CSS under different fading environments is introduced in section II. Section III briefly introduces the concept of GCN. Section IV develops GCN based framework for CSS. Extensive simulations and analyses are provided in Section V. Finally, section VI concludes this article.

II. MULTI-ANTENNA COOPERATIVE SPECTRUM SENSING UNDER FADING ENVIRONMENT

In this section, we consider a single−input−multiple−output (SIMO) cognitive radio network which consists of a single PU and S SUs, where the PU is equipped with a single antenna and the SUs are equipped with M antennas [21]. A SU performs spectrum sensing by using M antenna system and collects total N observations during a sensing period. The received PU signal at m-th antenna of s-th SU can be written as

$$y_s^m(n) = \begin{cases} h_s^m(n)x(n) + \eta_s^m(n), & \mathcal{H}_1 \\ \eta_s^m(n), & \mathcal{H}_0 \end{cases} \tag{1}$$

$n = 1, 2, 3, ..., N$, $s = 1, 2, ..., S$ and $m = 1, 2, ..., M$, where $x(n)$ indicates signal vector which can be assumed as independent and identically distributed (i.i.d) Gaussian vector with zero mean and variance σ_x^2 [8], [22]. In equation (1), $\eta_s^m(n)$ represents noise received at antenna m of SU s which is white Gaussian noise and is modeled as i.i.d Gaussian random vector with mean zero and variance σ_η^2. Considering the two states of PU, two hypotheses as $\mathcal{H}_0$ and $\mathcal{H}_1$ indicate the absence and presence of PU respectively. $h_s^m(n)$ represents the channel model of various kinds of fading present in wireless environment observed at antenna m of SU s. The channel gains of the PU–SU and SU–fusion centre channels are assumed to be constant during a sensing period, which can be justified by the flat-fading or narrowband channel over these links, where the sensing duration is shorter than the channel coherence time. During multiple sensing periods $h_s^m(n)$ follows a fading distribution [15], [22]. Fig. 1 depicts the multiple-SU with multiple-antenna CSS framework in a cognitive radio network. The fusion centre constructs the sensing matrix from individual sensing results collected from all SUs. To demonstrate the sensing performance, the probability of detection and probability of false alarm defined as $P_d = P[\mathcal{H}_1|\mathcal{H}_1]$ and $P_f = P[\mathcal{H}_1|\mathcal{H}_0]$ respectively.

A. Models for characterization of fading environment

To adapt to the dynamic changes in the wireless environment, we assumed that each SU is equipped with multiple antennas, signals received at the antennas have propagated through multiple paths and experience a different level of fading. Although there are various fading models which can be used for characterization of the signal propagation between the PU and the SU, in our work, we have considered three fundamental fading models named Rayleigh, Rician, and Weibull fading. The motivation for choosing these basic propagation models in our work is as follows. Each SU in the proposed system model is having multiple antennas spatially placed in such a way that some antennas on a SU may experience Rician fading having one dominating Line-Of-Sight (LOS) path along with other Non-Line-Of-Sight (NLOS) paths and other antennas may face Rayleigh/Weibull having only multiple NLOS paths fading with different values of fading parameters. All the three fading models are used to characterize the signal propagation in different practical indoor/outdoor environments.

Rayleigh fading is viewed as a suitable model for signal propagation in heavily built-up urban environments and is assumed to be most applicable when no dominant propagation path between transmitter and receiver is available. Empirical studies have shown that the Weibull fading model is used to characterize signal propagation in both indoor and outdoor environments [23]. The fading model for the Weibull distribution considers a signal composed of clusters of one multipath wave, each propagating in a non-homogeneous environment. The Rician fading is the stochastic model for radio wave propagation and is caused by partial cancellation of a radio signal by itself i.e. signals reaching to receiver have propagated through different paths exhibit multipath interference and at least one of the paths have shortened or increased in length. Rayleigh fading is considered a special case of Weibull fading and Rician fading. A brief description of these three fading channels is provided below:

1) Rayleigh fading channel: Rayleigh distribution is a continuous probability distribution for positive random variables. The probability density function (pdf) for the distribution is given by [24]

$$f_R(a) = \frac{a}{\sigma^2} exp\left(-\frac{a^2}{2\sigma^2}\right), \qquad a \geq 0 \tag{2}$$

where, a and σ^2 denote the envelope of Gaussian random variable X and Y with zero mean and variance respectively.

2) Rician fading channel: A multipath scenario considering the LOS path between transmitter and receiver is modeled using a rician fading channel. The pdf of received power in rician fading channel is defined as [24]

$$f_{Ri}(a) = \frac{a}{2\sigma^2} exp\left(-\frac{a^2 + a_0^2}{2\sigma^2}\right) I_0\left(\frac{aa_0}{2\sigma^2}\right), \tag{3}$$

where, a denotes the envelope of Gaussian random variables X with non zero mean a_0 and Y with zero mean. Parameter a_0 and σ^2 are defined in terms of k_0, $k_0 = a_0/2\sigma^2$. As the parameter k_0 increases, fading level decreases in the channel.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:56:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3214348

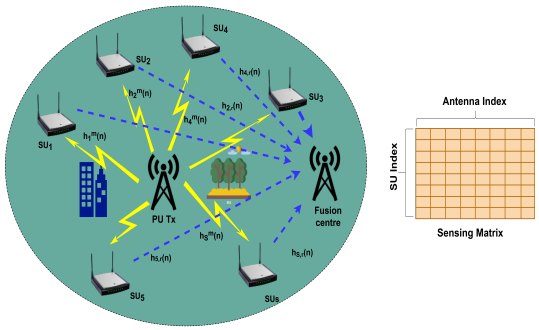


Fig. 1: Multi-antenna cognitive radio scenario

3) Weibull fading channel: The pdf of the power in the Weibull fading channel can be expressed as [24]

$$f_w(p) = \frac{m_w}{P_0} p^{m_w - 1} exp\left(-\frac{p^{m_w}}{P_0}\right), \tag{4}$$

where P_0 is average power equal to $2\sigma^2$. As the fading parameter m_w increases, fading level decreases.

We assume that each SU performs individual spectrum sensing which is based on energy detection for a sensing duration and the received signal strength (RSS) of the channel has to be accumulated. Consider T_s^m be the accumulated RSS at m-th antenna of s-th SU. Then T_s^m can be written as follows:

$$T_s^m = \frac{1}{N} \sum_{n=1}^{N} |y_s^m(n)|^2, \tag{5}$$

where, the channel between PU and m-th antenna of s-th SU, h_m^s in equation (1) can follow the fading model described in equations (2), (3) and (4) with varying fading severity. Consider $\mathcal{T}$ be the sensing matrix constructed from accumulated RSS at all M antennas of all S SUs, the sensing matrix in one time duration can be represented as

$$\mathcal{T} = \begin{bmatrix} T_1^1 & T_1^2 & T_1^3 & \cdots & T_1^M \\ T_2^1 & T_2^2 & T_2^3 & \cdots & T_2^M \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ T_S^1 & T_S^2 & T_S^3 & \cdots & T_S^M \end{bmatrix}. \tag{6}$$

The $\mathcal{T}$ matrix will be utilized at the fusion centre for the detection of PU state.

B. Modeling the Hidden node problem

Considering the practical scenarios, where some of the SUs may go far away from the PU/Fusion centre and they become hidden nodes and may not participate in the decision making. This is caused by many factors which include the large distance between the PU and SU and shadowing/multipath fading observed by the SU while sensing frequency bands occupied by the PU. In this work, we have addressed this problem under different fading channel conditions. To determine whether any of the SUs are hidden nodes or not, we have followed following assumptions

1) If the data/information is received from the SUs at the Fusion centre, we consider the SUs as active.
2) If no data/information is received from the SUs at the Fusion centre for a particular sensing period, we consider the SUs as hidden, as they are not contributing in the final decision making about the presence/absence of PU due to imperfect sensing and reporting channel.

Also, the number of hidden nodes may vary depending upon the SUs' mobility and channel severity faced by the users. Assume that ζ number of SUs are hidden which are not able to contribute to spectrum sensing. The updated sensing matrix considering the hidden nodes can be represented as:

$$\mathcal{T} = \begin{bmatrix} T_1^1 & T_1^2 & T_1^3 & \cdots & T_1^M \\ T_2^1 & T_2^2 & T_2^3 & \cdots & T_2^M \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ T_{S-\zeta}^1 & T_{S-\zeta}^2 & T_{S-\zeta}^3 & \cdots & T_{S-\zeta}^M \end{bmatrix} \tag{7}$$

where, $\zeta = (1, 2, 3, ..., \mathcal{Z})$, and $\mathcal{Z}$ is an integer, $\mathcal{Z} < S$.

C. Imperfect Reporting channel

The system models adopted by the existing DL based CSS methods consider fading and shadowing only in sensing channel under the assumption of perfect reporting channels, but practically it is impossible for fusion centre to receive error free sensing outcomes from SUs because of fading or shadowing.

The hidden node problem not only exist due to erroneous sensing channel but also can exits due to imperfect reporting channel as well. Therefore, we have considered the scenario of imperfect reporting channel while modeling the dynamic wireless environment.

Assuming no delay in forwarding local sensing outcomes to the FC, each SU forwards instantly its measured signal to the FC. The received signal at fusion centre from s-th SU $y_{s,c}(n)$ can be written as [25]

$$y_{s,c}(n) = \begin{cases} h_{s,r}(n)(h_s^m(n)x(n) + \eta_s^m(n)) + \eta_c(n), & \mathcal{H}_1 \\ h_{s,r}(n)\eta_s^m(n) + \eta_c(n), & \mathcal{H}_0 \end{cases} \tag{8}$$

$h_{s,r}(n)$ and $\eta_c(n)$ denotes the channel gain between s-th SU to fusion centre and zero-mean i.i.d Gaussian noise with variance σ_η^2. Each SU reports its sensing outcome to fusion centre through the reporting channel. In case of idle reporting channel, the received signal at the fusion centre can be denoted by equation (1), however, in case of erroneous reporting channel, the equation (1) can be modified as (8). The sensing outcomes of each SUs reported to fusion centre can experience small-scale fading. Therefore, we get the updated sensing matrix at fusion centre by using the equation (8) and (5). As a result of erroneous reporting channel, the detection performance of CSS is further decreased [18]. Therefore, it becomes necessary to consider the impact of imperfect reporting channel on the system performance. Due to severity of fading, sometimes some of SUs may report erroneous sensing outcomes to fusion centre or sometimes they become hidden nodes.

III. Fundamenatals of GCN

This section describes the basic preliminaries of GCN. The GCN generalizes the convolution operation from grid data

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:56:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3214348

to graph data. The basic principle behind graph convolution is to compute the node's features by aggregating its own features and neighbour's node features. GCN can be categorized into two methods: Spatial-based method and Spectral-based method. However, the GCN model proposed in [19] has bridged up the gap between spectral-based and spatial-based methods.

A. Graph representation

A graph, which is a non-linear data structure, represented in non-Euclidean space, describes complex relationships and inter-dependencies between objects. We assume $\mathcal{G}$ be an undirected graph defined as $\mathcal{G} = (\boldsymbol{V}, \boldsymbol{A}, \boldsymbol{E})$, where V is represented as set of nodes/vertices, $\boldsymbol{E}$ is represented as set of edges and $\boldsymbol{A}$ is adjacency matrix. Consider $v_i \in \boldsymbol{V}$ represents a node and if v_i points to v_j, then relationship between the nodes is denoted by $e_{ij} = (v_i, v_j) \in \boldsymbol{E}$.

B. Construction of adjacency matrix

Adjacency matrix can be defined as relationships between the nodes of the graph. The elements of adjacency matrix $\boldsymbol{A}$ can be calculated from the radial basis function (RBF):

$$\boldsymbol{A}_{i,j} = \exp\left(-\frac{\|x_i - x_j\|^2}{\rho^2}\right) \tag{9}$$

where x_i and x_j vectors represent the spectral signatures of nodes v_i and v_j respectively.

The parameter ρ controls the width of RBF. We can calculate Laplacian matrix from the matrix $\boldsymbol{A}$ as follows:

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A} \tag{10}$$

where $\boldsymbol{D}$ is diagonal matrix which denotes the degree matrix of $\boldsymbol{A}$, i.e., $\boldsymbol{D}_{i,i} = \sum_j \boldsymbol{A}_{i,j}$. The $\boldsymbol{L}_{sym}$ symmetric normalized Laplacian matrix can be defined as

$$\boldsymbol{L}_{sym} = \boldsymbol{I} - \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{L} \boldsymbol{D}^{-\frac{1}{2}} \tag{11}$$

where $\boldsymbol{I}$ denotes the identity matrix.

C. Graph convolution from Spectral perspective

Our work follows GCN in spectral domain as we employ a spectral-based approach in which spectral convolution on graphs is performed. Graphs are assumed to be undirected. Graph convolution in spectral domain are based on graph Fourier transforms. The Spectral based methods have a strong mathematical foundation in graph signal processing. Let us consider a graph signal $\boldsymbol{X} \in \mathcal{R}^{P \times F}$, where F denotes features (F features for each node), which is a feature vector of all nodes of a graph and $\boldsymbol{X}_i$ denotes feature of i-th node in a graph. By decomposing $\boldsymbol{X}$ in spectral domain, the spectral convolution between graph feature matrix $\boldsymbol{X}$ and a filter $g_\theta = \mathrm{diag}(\theta)$ parameterized by $\theta \in \mathcal{R}^P$ can be defined as

$$\mathcal{G}[\boldsymbol{X} * g_\theta] = U g_\theta U^T \boldsymbol{X} \tag{12}$$

where $\boldsymbol{U}$ denotes the matrix of eigenvectors of the matrix $\boldsymbol{L}_{sym}$. We can compute $\boldsymbol{U}$ as $\boldsymbol{L}_{sym} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^T$ and g_θ represents the function of eigenvalues (Λ) of $\boldsymbol{L}_{sym}$ with respect to variable θ, i.e., $g_\theta(\boldsymbol{A})$. $\boldsymbol{U}^T \boldsymbol{X}$ is the graph Fourier transform of feature matrix $\boldsymbol{X}$.

We need to calculate Laplacian eigenvectors for solving the equation (12), which is computationally complex for large graphs. To circumvent this issue, Hammond *et. al.* [26] proposed $K-$localized convolution on the graph by approximating the filter g_θ by applying Chebyshev polynomials up to the $K-$th order. By doing so, (12) can be written as

$$\mathcal{G}[\boldsymbol{X} * g_\theta] \approx \sum_{k=0}^{K} \theta_k^{'} T_k(\boldsymbol{L}_{norm}) \boldsymbol{X} \tag{13}$$

where $\theta_k^{'}$ and T_k are chebyshev coefficient and chebyshev polynomials repectively. The normalized Laplacian matrix $\boldsymbol{L}$ can be written as $\boldsymbol{L}_{norm} = (2/\lambda_{max})\boldsymbol{L}_{sym} - \boldsymbol{I}$. Here, λ_{max} is the largest eigenvalue of L_{sym}. The equation (13) is further simplified by Kipf and Welling [19] by limiting $K = 1$ and approximating the largest eigenvalue $\lambda_{max} \approx 2$ and can be rewritten as

$$\mathcal{G}[\boldsymbol{X} * g_\theta] \approx \boldsymbol{\theta}\left(\boldsymbol{I} + \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{A} \boldsymbol{D}^{-\frac{1}{2}}\right) \boldsymbol{X}. \tag{14}$$

By using eq. (14), consider a GCN with the following propagation rule

$$\boldsymbol{H}^{(l+1)} = \sigma(\boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{A} \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{H}^{(l)} \boldsymbol{W}^{(l)}) \tag{15}$$

where, $\boldsymbol{H}^{(l)}$ and $\boldsymbol{H}^{(l+1)}$ denotes input and output of lth GCN layer respectively, and $\boldsymbol{W}^{(l)}$ denotes the weights. $\sigma(\cdot)$ represents Rectified Linear Unit (ReLU) activation function.

IV. GCN based Cooperative Spectrum Sensing (GCN-CSS) framework

In this section, we provide a brief introduction to the construction of graph dataset and proposed GCN architecture for CSS. Here, we mainly focus on the construction of graphs from sensing matrices and perform reasoning on the graph signal. We use graph signal as the input to GCN which can extract the high-level features by aggregating the sensing information from the neighborhood of graph nodes. We enable the graph to be updated dynamically to adapt to the variations in the dimension of the sensing matrix due to the non-availability of input from hidden nodes. On the contrary, the CNN-based methods assume sensing matrices to be static, hence they are not able to model the dynamic changes that occur in the real environment. Therefore, we fed dynamic graphs (containing a variable number of nodes and dimension of adjacency matrices) as inputs to GCN for learning the features applicable for the spectrum state classification.

A. Graph data construction

As discussed in section III-C the graph feature matrix $\boldsymbol{X}$ is constructed from the sensing matrix. The graph feature matrix or graph signal $\boldsymbol{X}$ is consists of nodes' features on a graph. The graph feature $\boldsymbol{X}$ and adjacency matrix $\boldsymbol{A}$ is of dimensions $P \times F$ and $P \times P$ respectively. As it is known that $P = SM$ denotes total number of nodes in a graph, where each node represents a particular antenna equipped at particular SU and the feature

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:56:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3214348

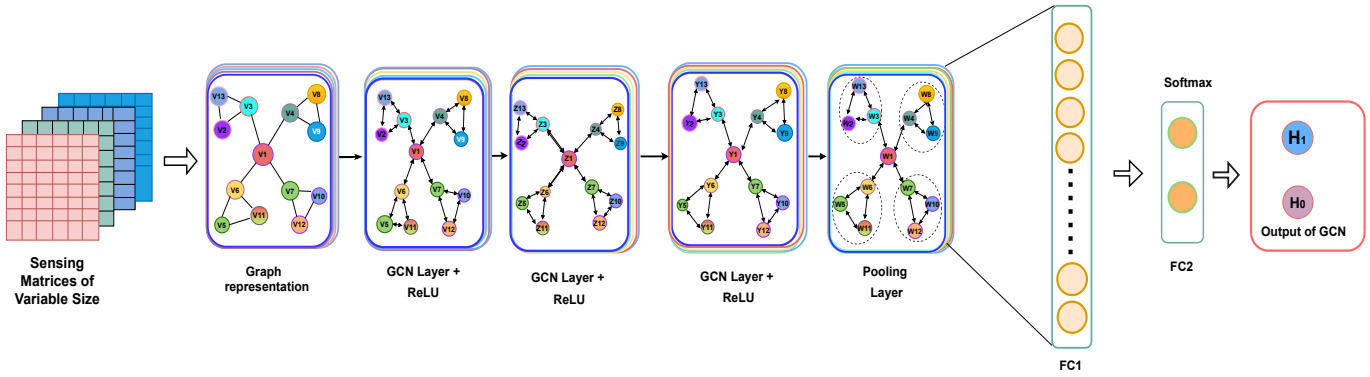


Fig. 2: The architecture of the proposed GCN-CSS method

value of each node is accumulated RSS received at the antenna of SU. F denotes number of features of a vertex/node, in our graph data, value of $F = 1$, which is RSS value.
Graph signal $\boldsymbol{X}$ is $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, ...\boldsymbol{X}_P]^T$, defined as per equation (6). where $\boldsymbol{X}_1 = T_1^1, \boldsymbol{X}_2 = T_1^2, \boldsymbol{X}_3 = T_1^3, ...\boldsymbol{X}_M = T_1^M, ..., \boldsymbol{X}_{M+1} = T_2^1, \boldsymbol{X}_{M+2} = T_2^2, \boldsymbol{X}_{M+3} = T_2^3, \boldsymbol{X}_{M+4} = T_2^4, ...\boldsymbol{X}_{2M} = T_2^M, ..., .., \boldsymbol{X}_P = T_S^M$. The value of P can vary due to variable number of hidden nodes ζ. In case, number of hidden nodes is ζ , then $P = (S - \zeta)M$. Each element of A is computed according to the equation (9), which is distance between a particular node to other node in the graph.

B. GCN-CSS architecture

Fig. 2 shows the proposed GCN architecture based spectrum sensing framework. The GCN-CSS architecture consists of graph convolutional layers, graph pooling layers, fully connected layer, and softmax layer. The graph convolutional layers extract high-level node representations, whereas graph pooling layers perform downsampling operations and coarsens the graph representations into subgraphs. The end-to-end framework of GCN for graph level classification is developed by applying a fully connected layer and Softmax layer to graph representations. The different layers in GCN architecture are described below.

1) Graph convolution layer: In CNN, conventional convolutions operate on regular, local grid structured data, while graph convolution operation allows every node to attend every other node on the graph. Graph convolution layer performs convolution of features from each node. In our case, the process of graph convolution is observed as message passing inside the whole graph. The output of graph convolutions layer is the convolved features of each node of a graph and denoted as:

$$\boldsymbol{Z}^{(l+1)} = \sigma(\boldsymbol{A}\boldsymbol{Z}^{(l)}\boldsymbol{W}^{(l+1)}) \tag{16}$$

where, $Z^{(0)} = X.W^{(l+1)}$ represents the learnable weights of a graph convolution layer, and $\sigma(\cdot)$ represents Rectified Linear Unit (ReLU) activation function. From equation (16), it is shown that a graph convolutional layer includes two steps: 1) generating new feature representation from the input $Z^{(l)}$ by performing a graph convolution, i.e. $\boldsymbol{A}\boldsymbol{X}\boldsymbol{Z}^{(l)}$ and 2) feeding the new generated features to a fully connected layer.

2) Graph pooling layer: Graph pooling layer does the same operation of reducing the number of nodes in a graph which is similar to pooling in conventional CNN. The pooling operation is performed in three steps, in first step the nodes of input graph are selected to form a group of nodes called supernode, subsequently, each supernode and its attributes are aggregated to form output node, and finally each output node is connected through edges and outputs a pooled graph. Therefore, pooling operation reduces the size of graph by aggregating or discarding the redundant information of the nodes so that the NN can learn a hierarchical representation of input data. This layer computes the coarser version of the graph and reduces the number of nodes in the graph at each step which results in single vector representations.

3) Fully connected layer: A multilayer perceptron network is fully connected (FC) to the pooling layer which consists of a hidden layer FC1 and output layer. To figure out the unique asset of graph convolutions, we compare them with fully connected layers, in which the layerwise forward propagation rule is

$$\boldsymbol{Z}^{(l+1)} = \sigma(\boldsymbol{Z}^{(l)}\boldsymbol{W}^{(l+1)}) \tag{17}$$

The difference between the forward propagation rule of fully connected layer and graph convolution layer is that of adjacency matrix applied on left of $\boldsymbol{Z}^{(l)}$. The benefit that this matrix brings is Laplacian smoothing, which calculates new features of a vertex as a weighted average of features of its neighbors on a graph. Provided that nodes in the same cluster are more likely to be densely connected, Laplacian smoothing allows them to have similar features, which makes the subsequent classifications much easier.

4) Classification: For classification task, we make use of softmax function on output features of the last FC layer of GCN, that is

$$\widehat{\boldsymbol{Y}}_{\boldsymbol{o}} = softmax(\boldsymbol{Z}^{l_{FC}-1}\boldsymbol{W}^{l_{FC}}) \tag{18}$$

where, l_{FC} is number of FC layers.

C. GCN based detection algorithm

Unlike CNN, the GCN feed graph shaped data into the network with an adjacency matrix that models the relationship

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:56:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3214348

between the nodes. Distance between the antennas placed at various SUs scattered over a geographical area is modeled by an adjacency matrix. The graph data and adjacency matrices are computed from sensing matrices just before the start of the training process. The labeled sensing matrices are collected for the construction of the dataset. The dataset can be written as:

$$(\boldsymbol{\psi}_{\mathcal{T}}, \mathbf{Z}) = \{(\mathcal{T}_1, z_1), (\mathcal{T}_2, z_2), \cdots, (\mathcal{T}_K, z_K)\}. \quad (19)$$

Here, $(\mathcal{T}_k, z_k)$ denotes the k-th $(k = 1, 2, 3, \cdots K)$ example of dataset $(\boldsymbol{\psi}_{\mathcal{T}}, \mathbf{Z})$ and the dimension of $\mathcal{T}$ changes as the number of hidden nodes ζ changes. $z \in \{0, 1\}$ denotes the label, where $z_k = 1$ and $z_k = 0$ represent the hypotheses of $\mathcal{H}_1$ and $\mathcal{H}_0$ respectively. Therefore, for the k-th data point of $(\mathcal{T}_k, z_k)$, the label can be converted to a one-hot vector.

$$z_k = \begin{cases} [1, 0]^T, & \mathcal{H}_1 \\ [0, 1]^T, & \mathcal{H}_0 \end{cases} \quad (20)$$

As the input to GCN is in form of graph data, hence after computation of graph and adjacency matrix, the dataset can be represented as:

$$(\boldsymbol{\Psi}_{\mathcal{G}}, Z) = \{(\boldsymbol{V_1}, \boldsymbol{A}_1, \boldsymbol{E}_1, z_1), (\boldsymbol{V}_2, \boldsymbol{A}_2, \boldsymbol{E}_2, z_2), \cdots, (\boldsymbol{V}_K, \boldsymbol{A}_K, \boldsymbol{E}_K, z_K)\}. \quad (21)$$

Here, graph $\mathcal{G}$ is made up of $(\boldsymbol{V}, \boldsymbol{A}, \boldsymbol{E})$ and Z is the set of label z.

$$(\boldsymbol{\Psi}_{\mathcal{G}}, Z) = \{(\boldsymbol{\mathcal{G}_1}, z_1), (\boldsymbol{\mathcal{G}}_2, z_2), \cdots, (\boldsymbol{\mathcal{G}}_K, z_K)\}. \quad (22)$$

The labeled data points are fed to the GCN, which consists of three convolutional layer and one pooling layer. The output of last layer is flattened and fed to FC layers. Finally, the output of GCN is a 2x1 class score vector. We adopt softmax function to normalize the score vector and obtained the result as:

$$h_\theta((\boldsymbol{\mathcal{G}_k})) = \begin{bmatrix} h_{\theta|\mathcal{H}_1}(\boldsymbol{\mathcal{G}_k}) \\ h_{\theta|\mathcal{H}_0}(\boldsymbol{\mathcal{G}_k}) \end{bmatrix}, \quad (23)$$

with

$$h_{\theta|\mathcal{H}_1}(\boldsymbol{\mathcal{G}_k}) + h_{\theta|\mathcal{H}_0}(\boldsymbol{\mathcal{G}_k}) = 1 \quad (24)$$

where $h_\theta(\boldsymbol{\mathcal{G}_k})$ is the expression of GCN network with the input graph and adjacency matix $(\boldsymbol{\mathcal{G}_k})$, and model parameter θ. $h_\theta/\mathcal{H}_i(\boldsymbol{\mathcal{G}_k})$ represents the class probability of hypothesis $\mathcal{H}_i$. The goal of model training is to maximize the likelihood

$$E(\theta) = \prod_{k=1}^{K} (h_{\theta|\mathcal{H}_1}(\boldsymbol{\mathcal{G}_k}))^{z_k} (h_{\theta|\mathcal{H}_0}(\boldsymbol{\mathcal{G}_k}))^{1-z_k}, \quad (25)$$

and the log likelihood

$$\begin{aligned} l(\theta) &= \log E(\theta) \\ &= \sum_{k=1}^{K} z_k \log h_{\theta|\mathcal{H}_1}(\mathcal{G}_k) + (1 - z_k) \log(1 - h_{\theta|\mathcal{H}_0}(\boldsymbol{\mathcal{G}_k})). \end{aligned} \quad (26)$$

Our main goal is to obtain the optimal parameter θ from the NN training and can be expressed as

$$\theta^* = \arg\max_{\theta} E(\theta). \quad (27)$$

To get the optimal parameter, we minimize the cross-entropy loss function and use the Adam optimizer . Using the back-propagation algorithm, we update the model parameter and get a well-trained GCN model. Therefore, the outputs of well-trained GCN are $h_{\theta^*|\mathcal{H}_1}(\mathcal{G})$ and $h_{\theta^*|\mathcal{H}_0}(\mathcal{G})$. According to Neyman-Pearson's theorem, the optimal test statistics is the likelihood ratio [27]. Hence, GCN based test statistic can be defined as

$$\beta_{GCN} = \frac{h^*_{\theta|\mathcal{H}_1}(\boldsymbol{\mathcal{G}_k})}{h^*_{\theta|\mathcal{H}_0}(\boldsymbol{\mathcal{G}_k})}. \quad (28)$$

We can decide the state of PU for test data by comparing the output of GCN. However, this method can not control the (P_f). Therefore, we use the Monte-Carlo method to determine a detection threshold γ to achieve the desired false alarm probability α. Let

$$\boldsymbol{\Psi}_{\mathcal{G}_{L_n}} = \{(\boldsymbol{\mathcal{G}_1}, 0), (\boldsymbol{\mathcal{G}}_2, 0), \cdots, (\boldsymbol{\mathcal{G}}_{L_n}, 0)\} \quad (29)$$

$\boldsymbol{\Psi}_{\mathcal{G}_{L_n}}$ denotes the dataset under $\mathcal{H}_0$ hypotheses, which consists of L_n examples. Therefore, we fed the graph data samples of noise to GCN and get the expression of test statistics under $\mathcal{H}_0$ i.e. $\beta_{GCN|\mathcal{H}_0}$. We re-arrange all the $\beta_{GCN|\mathcal{H}_0}$ in descending order and construct a set from them as denoted by $\boldsymbol{\Psi}_{\boldsymbol{\beta_{GCN|\mathcal{H}_0}}}$. Now, we can obtain the detection threshold with α constraint as follows:

$$\gamma = \boldsymbol{\Psi}_{\beta_{GCN|\mathcal{H}_0}(round((1-\alpha)L_n))}, \quad (30)$$

where, $round(\cdot)$ represents the round function which rounds a number to the nearest integer value. After calculating the detection threshold, for the new incoming test data, we can make decision about the state of PU by

$$\gamma \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \beta_{GCN}. \quad (31)$$

V. Complexity Analysis

A. Computational Complexity of CNN

Despite the outstanding results obtained by CNN based architectures, these models rely on 2-D convolutional layer and fully connected layers, where lth layer defines a kernel as data tensor $\boldsymbol{W}^{(l)}_{CNN} \in \mathcal{R}^{n^{(l)}_{ks} \times n^{(l)}_{ks} \times n^{(l)}_f \times n^{(l-1)}_f}$ is made up of $n^{(l)}_f$ filters with spatial size $n^{(l)}_{ks} \times n^{(l)}_{ks}$. Each convolutional layer involves $(n^{(l)}_{ks} \cdot n^{(l)}_{ks}) + 1) n^{(l)}_f$ parameters at least. The computational complexity of convolutiopnal layer is provided as [28]

$$\mathrm{O}\left(\sum_{l=1}^{L_{CNN}} n^{(l-1)}_f \cdot n^{(l)}_{ks} \cdot n^{(l)}_{ks} \cdot n^{(l)}_f \cdot m^{(l)} \cdot m^{(l)} \right) \quad (32)$$

Here l denotes the index of current convolutional layer, and L_{CNN} is number of convolution layers (i.e. depth of the model), $n^{(l)}_f$ and $n^{(l-1)}_f$ are the number of filters of the lth and $(l-1)$th layers, respectively. n_{ks} is spatial size of the current layer, and m^l is spatial size of resulting output feature map. In this sense, it is observed that spatial convolutions $n_{ks} \times n_{ks}$ are expensive, increasing both computational time and the models size quadratically with respect to n_{ks}.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:56:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3214348

B. Computational Complexity of GCN

For the complexity of the information aggregation in (16), $X^{(l)} = \boldsymbol{AXZ}^{(l)}$ costs $\mathrm{O}(|\boldsymbol{E}|\boldsymbol{F})$, and $X^{(l)}\boldsymbol{W}^{(l)}$ costs $\mathrm{O}(PF^2)$. Hence, the complexity of L-layer GCN is provided by following equation [29]

$$\mathrm{O}\left(L_{GCN}|E|F + L_{GCN}PF^2\right) \tag{33}$$

where, L_{GCN} is the number of graph convolution layers in GCN, $|E|$ is number of edges.

C. Computational complexity of Artificial Neural Network

The ANN architecture consists of multiple FC layers. The computational complexity of ANN provided as $\mathrm{O}\left(\sum_{l=1}^{L_{ANN}} n_d^{(l)} n_d^{(l-1)}\right)$, where L_{ANN} is the number of FC layers in ANN, $n_d^{(l)}$ and $n_d^{(l-1)}$ are the number of neurons in lth and $(l-1)$th FC layers, respectively.

D. Computational complexity of SVM and K-means clustering

Consider that input feature vector fed to SVM is of length D_{in} then the complexity for SVM can be estimated as [30] $\mathrm{O}(smD_{in})$. Where, s is number of iterations and m is total number of training points. The computational complexity of K-means algorithm can be computed as [31] $\mathrm{O}(mN_cD_{in})$, where N_c is number of clusters.

VI. Numerical Results

A. Sensing Data Generation

We consider that the PU signals are i.i.d signals in the presence of Gaussian noise, and the noise on each antenna of SU is assumed to follow i.i.d. zero-mean Gaussian distribution with variance σ_η^2. We have arbitrarily considered 15 SUs ($S = 15$), where each SU is equipped with 15 antennas ($M = 15$) and N equal to 500. We have adapted the multi-antenna scenario at a SU based on [8] and [21], where the effect of variation in number of antennas at a SU and the number of SUs on the detection probability is extensively studied. Hence, increasing the number of antennas at a SU would improve the spatial diversity and spatial multiplexing gain, hence improving the detection performance. The developed model is valid for any number of S and M , however for choosing the optimal number of SUs based on the available bandwidth of the reporting channel, the readers may refer [32]. The sensing matrices are constructed from accumulated RSS at all antennas of each SU and it is assumed that at least one antenna of each SU can experience Rician fading and other antennas may experience Rayleigh/Weibull fading with different values of fading parameters. We considered a range of fading parameters for three different fading channels. The amount of fading (AF) is a measure of severity of a fading channel. The AF depends on a single parameter and fading parameter values corresponding to three different fading channels are provided in Table I. The parameter a in equation (2), parameter a_0 in equation (3) and parameter m_w in equation (4) are varied to change the fading severity. We have simulated the dataset by making an assumption that a PU signal reaching to multiple antennas of a SU can experience different type of fading with different amount of fading parameter considering perfect reporting channel. The developed model does not have any limitation on the type of fading experienced by the sensing signal. We have made the process of choosing a fading channel and fading parameter is completely random for the antennas of SUs i.e. the sensing signal can undergo any type of fading with different level of fading severity before reaching the antenna. Hence, dataset construction procedure ensures that dynamic wireless environment conditions have been taken into account. The sensing matrices can be computed from equation (6) and converted to graph data as discussed in section IV-A.

TABLE I: Fading parameter values for different channels

Fading channel	**Fading parameter values**				
Rayleigh (a)	0.58	0.6	0.61	0.62	0.65
Rician (k_0)	0	0.3	0.4	0.5	0.55
Weibull (m_w)	2	2.5	3	3.5	4

B. Simulation Environment and Hyperparameters

In simulations, we have utilized the Spektral library [33] for building graphs, graph convolutional networks as well as Keras library with TensorFlow backend for creating and training the NN models. Table II shows the hyperparameters along with training parameters used in simulations during the training of the GCN-CSS, where "ReLU" and "Softmax" represent the activation functions, and number of learnable parameters. The hyperparameters of CNN and learnable parameters are depicted in Table III. The ANN model consists of two fully connected layer with number of neurons as 100 and 2 respectively.

Table II also highlights the dimension of graph data used for simulation based on arbitrarily chosen number of hidden nodes. Considering the number of hidden nodes $\zeta = 0, 2$ and 4, the dimension of vertices of graph data will be (225×1), (195×1) and (165×1) respectively. The proposed GCN-CSS model is trained on a single dataset which consists of all these three different types of variable size graphs corresponding to different number of hidden nodes.

Since, CNN expects input matrix of fixed dimension, therefore we fill the elements with zeros in the row of sensing matrix representing the sensing signal of the hidden node. The input for the CNN model consists of three types of sensing matrices each corresponding to the number of hidden nodes $\zeta = 0, 2$ and 4 and the dimension of the matrix is : 15×15 . For other ML models, the same methodology is applied to cater the hidden nodes in the data.

C. Simulation results under perfect Reporting Channel

We evaluate the detection performance in terms of Receiver Operating Characteristics (ROC) curve, (P_d) versus SNR through comparing the proposed method with CNN, ANN, SVM, and K-Means clustering methods.

1) Detection performance evaluation: In this section, we have evaluated the performance of GCN-CSS method by considering the imperfect sensing channel and perfect reporting channel.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:56:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3214348

TABLE II: Hyperparameters of GCN-CSS

Input:			
Graph 1: $\boldsymbol{X}$ (Dimension: 225 × 1), $\boldsymbol{A}$ (Dimension: 225 × 1)			
Graph 2: $\boldsymbol{X}$ (Dimension: 165 × 1), $\boldsymbol{A}$ (Dimension: 165 × 165)			
Graph 3 : $\boldsymbol{X}$ (Dimension: 195 × 1), $\boldsymbol{A}$ (Dimension: 195 × 195)			
Layers	No. of Units in Conv layer	No. of Input feature	Learnable Parameters
GCNConv1 + ReLU	32	1	64
GCNConv2 + ReLU	64	32	2112
GCNConv3 + ReLU	128	64	8320
Global Sum Pooling			
FC 1 + ReLU	64	128	8,256
FC 2 + Softmax	2	64	130
Class Score Vector	(Dimension: 2 × 1)		
Total Parameters			18,882

TABLE III: Hyperparameters of CNN

Input: ($\mathcal{T}$) : (Dimension : 15 ×15), for (ζ = 0, 2 and 4)				
Layers	Kernel size	Input shape	Onput shape	Learnable Parameters
Conv1 + ReLU	5x5x32	(15x15x1)	(15x15x32)	832
Max-pool 1	2x2	(15x15x32)	(7x7x32)	-
Conv2 + ReLU	3x3x64	(7x7x32)	(7x7x64)	18,496
Max-pool 2	2x2	(7x7x64)	(3x3x64)	-
Conv3 + ReLU	3x3x128	(3x3x64)	(3x3x128)	73,856
Max-pool 3	2x2	(3x3x128)	(1x1x128)	-
FC 1 + ReLU	-	128	128	16,512
FC 2 + Softmax	-	128	2	258
Total Parameters				109,954

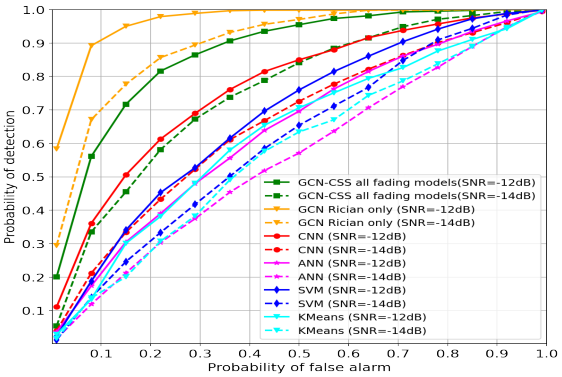


Fig. 3: ROC curve for different CSS methods

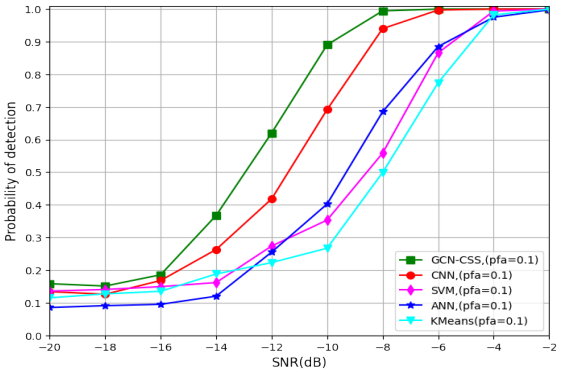


Fig. 4: Pd vs SNR plot at $P_f = 0.1$

To evaluate the detection performance, we calculate the detection threshold at different P_f values to obtain the corresponding P_d values. We have considered the number of hidden nodes as 2 and 4 for the proposed GCN-CSS method. Fig. 3 shows the ROC curve for comparing the performance of GCN-CSS method with other benchmark methods at two different values of SNR i.e. −12dB and −14dB. To test the robustness of the proposed model towards different fading environment, we have simulated two different scenarios. In the first scenario, only single type of fading is considered i.e. Rician, which implies that all the antennas of a SU are receiving the sensing signal that has experienced Rician fading. In the second scenario, all the three types of fading channels i.e. Rician, Weibull and Rayleigh are considered. Fig. 3 clearly highlight that the performance of the GCN-CSS model where only single type of fading is considered performs better than the second scenario. From the results, we can observe that the detection performance of the proposed GCN-CSS method is higher as compared to other methods like CNN, ANN, SVM, and K-Means method.

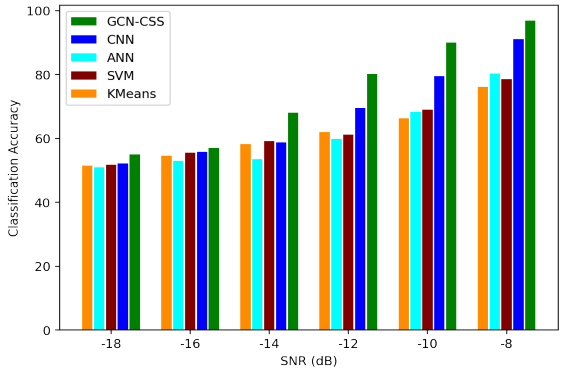


Fig. 5: Classification Accuracy vs SNR

To further evaluate the detection performance, a plot of P_d versus SNR at $P_f = 0.1$ is shown in Fig. 4. According to the IEEE 802.22 standard, the value of false alarm is set as ($P_f = 0.1$). It can be seen that the detection performance is highest for the proposed GCN-CSS method as compared to other methods. This is because the GCN-CSS is able to model the dynamic changes in the real environment, whereas the other methods fail to do so. Fig. 5 shows the comparison of classification accuracy of the proposed method with other methods for different SNR values. From the bar plot, we can observe that the proposed method achieves better classification accuracy as compared to different algorithms.

2) Ablation study for Hidden Nodes: To show the effect of hidden nodes on the sensing performance of the proposed method, we have performed an ablation study where the number of hidden nodes is varied and the consequences are evaluated. To prove the performance of GCN-CSS superior in practical scenario, we have shown the effect of the number of hidden nodes on P_d and simulations are carried out by considering $0, 4, 6$ and 8 number of hidden nodes. Fig. 6 depicts the ROC curve for varying values of hidden nodes and in Fig. 7 the bar graph shows the effect of the number of hidden nodes on P_d for various values of SNR. From the results, it is observed that as the number of hidden nodes increases, detection probability slightly decreases. However,

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:56:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3214348

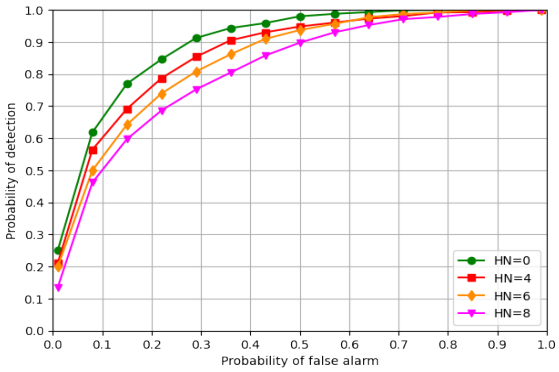


Fig. 6: ROC curve showing effect of hidden nodes (SNR=-12dB)

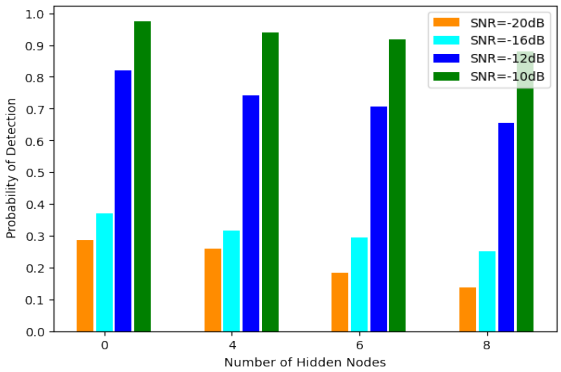


Fig. 7: Bar plot showing the effect of hidden nodes on P_d for different SNR

GCN is still able to perform better in the case where hidden nodes are present.

D. Simulation results under Imperfect Reporting channel

We have evaluated the performance of GCN-CSS considering imperfect sensing channel and imperfect reporting channel. The impact of imperfect reporting channel is analysed by considering the effect of any type of fading with different severity on SUs sensing outcomes. For generating data, we consider the fading environment for reporting channel similar to the fading environment assumed in case of sensing channel, as described in section II-A.

When sensing signals of SUs transmitted through reporting channel, the received signal at fusion centre from s-th SU can be defined by equation (8). The SUs sensing outcomes reaching to fusion centre are subjected to any type of the fading channel and fading parameter as described by equation (2), (3), and (4), and as per Table I respectively. The effect of imperfect reporting channel on detection performance of GCN-CSS and CNN based CSS method due to fading is depicted in figure 8. It is clearly evident that the imperfect reporting channel has deteriorated the detection performance at fusion centre. However, the proposed GCN-CSS method shows better detection performance in comparison with the CNN based method even in the case of imperfect reporting channel.

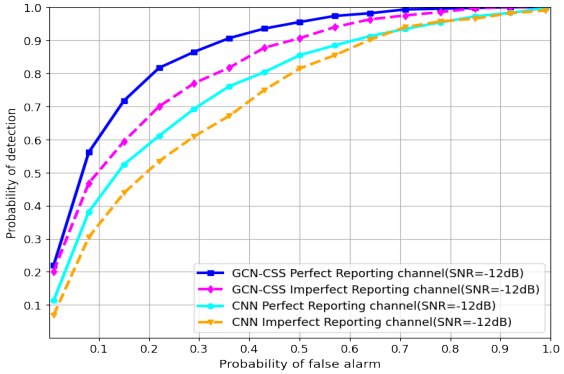


Fig. 8: ROC curve showing effect of Imperfect Reporting channel

E. Comparison in terms of complexity

We have reported the complexity of convolutional layers of both GCN-CSS and CNN methods as the computations required in fully connected layers are equal for both the GCN and CNN architectures. As shown in Table II and III, the learnable parameters are $18,882$ and $109,954$ for GCN-CSS and CNN method respectively, whereas the learnable parameters of ANN are $22,802$. The average inference time (ms/sample) using 100 data samples is calculated for GCN-CSS, CNN and ANN method as 0.010, 0.037 and 0.034 respectively. The inference time for proposed GCN-CSS method is the summation of computation time taken to compute the Graph and adjacency matrix and model inference time for one data sample. From the analysis, it is concluded that the proposed GCN-CSS method has less complexity since the required number of learnable parameters and inference time are less as compared to CNN and ANN. Further, the simulation results conclude that the proposed GCN-CSS method provides optimal detection performance even under the severe fading scenarios as well as in the presence of hidden nodes.

VII. Conclusion

In this work, we have proposed a CSS scheme utilizing GCN which is capable of adapting the dynamic changes in the wireless environment. To this end, learning CNN on the static sensing matrix does not provide optimal performance under practical scenarios. Hence, it becomes important to study the efficient dynamic graph convolutional network models. As the extension of CNN, GCN is capable of handling the non-grid data, thus it can easily adapt in the practical scenarios of one or more hidden nodes. The proposed methodology acquires better sensing performance in comparison with the standard algorithms like CNN, ANN, SVM, and K-Means clustering. The proposed GCN-CSS method is less computationally complex as compared with CNN based method and also provides better sensing performance even in the case of imperfect reporting channel. It is also concluded, that the performance gain of our solution in comparison to the other algorithms is optimal in case of an increase in the number of hidden nodes.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:56:30 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3214348

REFERENCES

[1] S. Kumar, "Performance of ed based spectrum sensing over α–η–μ fading channel," *Wireless Personal Communications*, vol. 100, no. 4, pp. 1845–1857, 2018.

[2] S. Kumar, P. S. Chauhan, P. Raghuwanshi, and M. Kaur, "Ed performance over α-η-μ/ig and α-κ-μ/ig generalized fading channels with diversity reception and cooperative sensing: A unified approach," *AEU-International Journal of Electronics and Communications*, vol. 97, pp. 273–279, 2018.

[3] D. Janu, K. Singh, and S. Kumar, "Machine learning for cooperative spectrum sensing and sharing: A survey," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 1, p. e4352, 2022.

[4] O. P. Awe, A. Deligiannis, and S. Lambotharan, "Spatio-temporal spectrum sensing in cognitive radio networks using beamformer-aided svm algorithms," *IEEE Access*, vol. 6, pp. 25 377–25 388, 2018.

[5] J. Zhuang, Y. Wang, S. Zhang, P. Wan, and C. Sun, "A multi-antenna spectrum sensing scheme based on main information extraction and genetic algorithm clustering," *IEEE Access*, vol. 7, pp. 119 620–119 630, 2019.

[6] Y. Xu, P. Cheng, Z. Chen, Y. Li, and B. Vucetic, "Mobile collaborative spectrum sensing for heterogeneous networks: A bayesian machine learning approach," *IEEE Transactions on Signal Processing*, vol. 66, no. 21, pp. 5634–5647, 2018.

[7] C. Liu, X. Liu, and Y.-C. Liang, "Deep cnn for spectrum sensing in cognitive radio," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–6.

[8] C. Liu, J. Wang, X. Liu, and Y.-C. Liang, "Deep cm-cnn for spectrum sensing in cognitive radio," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2306–2321, 2019.

[9] Z. Chen, Y.-Q. Xu, H. Wang, and D. Guo, "Deep stft-cnn for spectrum sensing in cognitive radio," *IEEE Communications Letters*, vol. 25, no. 3, pp. 864–868, 2020.

[10] S. Zheng, S. Chen, P. Qi, H. Zhou, and X. Yang, "Spectrum sensing based on deep learning classification for cognitive radios," *China Communications*, vol. 17, no. 2, pp. 138–148, 2020.

[11] J. Xie, C. Liu, Y.-C. Liang, and J. Fang, "Activity pattern aware spectrum sensing: A cnn-based deep learning approach," *IEEE Communications Letters*, vol. 23, no. 6, pp. 1025–1028, 2019.

[12] N. Balwani, D. K. Patel, B. Soni, and M. Lopez-Benitez, "Long short-term memory based spectrum sensing scheme for cognitive radio," in *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2019, pp. 1–6.

[13] B. Soni, D. K. Patel, and M. Lopez-Benitez, "Long short-term memory based spectrum sensing scheme for cognitive radio using primary activity statistics," *IEEE Access*, vol. 8, pp. 97 437–97 451, 2020.

[14] L. Yu, J. Chen, G. Ding, Y. Tu, J. Yang, and J. Sun, "Spectrum prediction based on taguchi method in deep learning with long short-term memory," *IEEE Access*, vol. 6, pp. 45 923–45 933, 2018.

[15] J. Xie, J. Fang, C. Liu, and X. Li, "Deep learning-based spectrum sensing in cognitive radio: A cnn-lstm approach," *IEEE Communications Letters*, vol. 24, no. 10, pp. 2196–2200, 2020.

[16] K. Yang, Z. Huang, X. Wang, and X. Li, "A blind spectrum sensing method based on deep learning," *Sensors*, vol. 19, no. 10, p. 2270, 2019.

[17] D. Ke, Z. Huang, X. Wang, and X. Li, "Blind detection techniques for non-cooperative communication signals based on deep learning," *IEEE Access*, vol. 7, pp. 89 218–89 225, 2019.

[18] M. Li, O. Alhussein, P. C. Sofotasios, S. Muhaidat, P. D. Yoo, J. Liang, and A. Wang, "Censor-based cooperative multi-antenna spectrum sensing with imperfect reporting channels," *IEEE Transactions on Sustainable Computing*, vol. 5, no. 1, pp. 48–60, 2019.

[19] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[20] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5966–5978, 2020.

[21] Y. Zhang, S. Zhang, Y. Wang, J. Zhuang, and P. Wan, "Riemannian mean shift-based data fusion scheme for multi-antenna cooperative spectrum sensing," *IEEE Transactions on Cognitive Communications and Networking*, 2021.

[22] L. Wei and O. Tirkkonen, "Spectrum sensing in the presence of multiple primary users," *IEEE Transactions on Communications*, vol. 60, no. 5, pp. 1268–1277, 2012.

[23] R. Singh, S. K. Soni, R. S. Raw, and S. Kumar, "A new approximate closed-form distribution and performance analysis of a composite weibull/log-normal fading channel," *Wireless Personal Communications*, vol. 92, no. 3, pp. 883–900, 2017.

[24] P. M. Shankar, *Fading and shadowing in wireless systems*. Springer, 2017.

[25] M. B. Ghorbel, H. Nam, and M.-S. Alouini, "Soft cooperative spectrum sensing performance under imperfect and non identical reporting channels," *IEEE Communications Letters*, vol. 19, no. 2, pp. 227–230, 2014.

[26] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[27] S. M. Kay, "Fundamentals of statistical signal processing. detection theory, volume ii," *Printice Hall PTR*, pp. 1545–5971, 1998.

[28] M. E. Paoletti, J. M. Haut, X. Tao, J. Plaza, and A. Plaza, "Flop-reduction through memory allocations within cnn for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5938–5952, 2020.

[29] Z. Zhang, C. Chen, Y. Chang, W. Hu, Z. Zheng, Y. Zhou, and L. Sun, "Hin2grid: A disentangled cnn-based framework for heterogeneous network learning," *Expert Systems with Applications*, vol. 187, p. 115823, 2022.

[30] C. Hou, F. Nie, C. Zhang, D. Yi, and Y. Wu, "Multiple rank multi-linear svm for matrix data classification," *Pattern Recognition*, vol. 47, no. 1, pp. 454–469, 2014.

[31] K. P. Sinaga and M.-S. Yang, "Unsupervised k-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80 716–80 727, 2020.

[32] Y. He, J. Xue, T. Ratnarajah, M. Sellathurai, and F. Khan, "On the performance of cooperative spectrum sensing in random cognitive radio networks," *IEEE Systems Journal*, vol. 12, no. 1, pp. 881–892, 2016.

[33] D. Grattarola and C. Alippi, "Graph neural networks in tensorflow and keras with spektral," 2020.



Dimpal Janu Dimpal Janu received her M.Tech. degree in Electronics and Communication Engineering from Malaviya national institute of technology (MNIT), Jaipur, India in 2018. She is currently pursuing a Ph.D. degree from MNIT Jaipur. Her research interests include wireless communication, cognitive radio network, machine learning, and deep learning.



Sandeep Kumar received his Master of Engineering in Electronics and Communication from Thapar University, Patiala, India in 2007. He received his Ph.D. from Delhi Technological University, Delhi, India in 2018. He is currently working as Member (Senior Research Staff) at Central Research Laboratory, Bharat Electronics Limited Ghaziabad, India. His research interests include the study of wireless channels, performance modeling of fading channels, and cognitive radio networks.



Kuldeep Singh received his MTech degree in Signal Processing from Delhi University in 2006 and a Ph.D. degree in Computer Vision from Delhi Technological University, India, in 2016. He is currently an assistant professor with the Department of Electronics & Communication Engineering, Malaviya National Institute of Technology, Jaipur, India. Previously, he was a Senior Scientist with the Central Research Lab, Bharat Electronics Ltd., India. His research interest includes Computer Vision, Machine/ Deep Learning, and Biometrics.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:56:30 UTC from IEEE Xplore. Restrictions apply.

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.

A Novel Recommender System Using PSO, FCM and, Autoencoder

Om Prakash Verma
Delhi Technological University
Nidhi Beniwal (nidhidtu15@gmail.com)
Delhi Technological University

Research Article

Keywords: Deep Learning, Recommender system, Fuzzy C-means Clustering, Autoencoder, Particle Swarm Optimization, Autoencoder

Posted Date: October 10th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2125860/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

A NOVEL RECOMMENDER SYSTEM USING PSO, FCM AND, AUTOENCODER

Om Prakash Verma[1][0000-0002-0860-9166] and Nidhi Beniwal[2][0000-0002-5791-5899]

[1] Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India

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

Email ID - nidhidtu15@gmail.com

ABSTRACT. The recommender system is the most developing application in the web environment. A good Recommendation system needs to make choices and decisions to provide recommendations to the user. The Recommender system filters the information and provides accurate recommendations of products or services for a concerned individual. In this research, we propose a model, which employs three approaches; clustering, nature-inspired optimizer, and deep learning. The fuzzy c-means clustering technique is combined with the particle swarm optimization algorithm. The particle swarm optimizer works as a feature extractor. The particle swarm optimizer is applied to the Movielens dataset to get the initial cluster positions. The fuzzy c-means is used to classify the users in the dataset and to reduce data redundancy. The optimized and clustered data is provided as input to the autoencoder to get the final recommendations of the movies to the users. We have proposed our recommender system in two parts; 1) Using a single-layer autoencoder, and 2) Using a deep autoencoder. We analyzed our proposed system over a publicly available dataset- Movielens. The proposed system is compared with the existing recommender systems. The experimental results of our system revealed that the performance and efficiency are improved, and offered better recommendations.

Keywords: Deep Learning, Recommender system, Fuzzy C-means Clustering, Autoencoder, Particle Swarm Optimization, Autoencoder

1. Introduction

Recommender System is a system that filters the information to provide personalized recommendations to the users. A recommender system is a software that suggests products to the users according to their interests or likes in the e-commerce domain. The recommender system has been of great use in the field of entertainment, online shopping, tourism, and other areas. In the online environment, the data is big, voluminous, complex, and dynamic. So, it solves the issues of information

overloading. The recommender system provides recommendations based on the user or item **[1]**. Recommender System are the lifesavers in the infinitely deep pool of information in the domain of e-commerce, by improving the customer experience. Recommender System helps in removing the injustice of choices, making a way for smooth decisions, and also boosting sales in the e-commerce domain**[2]**. In light of the COVID-19 pandemic, online sales have increased tremendously. There is competition among the businesses. So, businesses must ensure a good customer experience by providing better recommendations. Recommender Systems are mainly classified into three broad categories namely; 1) Content-based Recommender Systems, 2) Collaborative Filtering Recommender Systems and 3) Hybrid Recommender Systems**[3]**. The Content-Based recommender system gives recommendations based on a comparison of user's and item's additional information such as images, texts, and videos. Collaborative Filtering(CF) based recommender system is the most widely used system. The recommendations of Collaborative Filtering include learning from user-item interactions. The interaction data could be implicit(browsing history or cookies) or explicit(based on users ratings). Collaborative Filtering models can be further divided into two types- memory-based or model-based. The hybrid recommender system refers to integrating two or more types of recommendation models. Despite the better efficiency of the Collaborative Filtering-based model, the technique suffers from various problems such as cold-start, sparsity of data, and incapable of capturing complex interactions among users and items **[4]**. There are various bio-inspired optimizers such as Genetic Algorithm(GA), Ant Colony Optimization(ACO), Cuckoo Search, Artificial Bee Colony(ABC), and Fish swarm Algorithm(FSA)**[5][6]**. In our work, the Particle Swarm Optimization(PSO) algorithm**[7]** is used to get the initial cluster positions. PSO is a nature-inspired algorithm for feature reduction. Clustering is a technique used to group the data according to some criteria. K-means clustering and Fuzzy C-means clustering are the most widely used clustering techniques for recommender systems. Fuzzy C-means(FCM) is used in our work to get the required clusters, in which a user doesn't belong to a single cluster and it gives more accurate recommendations. Also, we have employed a deep learning-based algorithm- Autoencoder. Deep Learning is being used significantly in various areas such as image analysis, natural language processing, and pattern and speech recognition **[8]**. Recent studies also suggest the effectiveness of deep learning fo recommendation purpose. Due to deep learning's high performance and providing a better quality of recommendations, it has been widely used in recommender systems**[9]**. In comparison with traditional recommender systems, deep learning gives a better perception of the demands of the user, the characteristics of items, and the interactions between them **[10]**. Deep Learning imitates the human brain for interpretation of the data which can be images, texts, or sounds. Deep Learning is learning the deeper representations which means representations of multiple levels. Deep Learning originates from Artificial Neural Networks and it is the combination of various areas such as artificial intelligence, neural networks, Pattern Recognition, etc **[11]**. The structure of Deep Learning is a Multilayer Perceptron(MLP) which is

a feed-forward neural network with multiple levels of hidden layers. As deep learning is getting popular these days, there are some reasons behind that. The increased processing capability of the chip, the increase in the size of the training data, known as BIG DATA, and lastly, the advanced Machine Learning research are the main reasons to introduce Deep Learning. These have made the use of deep learning methods to be used for the complex and composite nonlinear functions **[12]**. It is also used to learn the feature representations which are hierarchical and distributed. It also makes effective use of both types of data, whether labelled or unlabelled **[13]**. The reason behind the use of deep learning is that it can make use of complex and non-linear relationships between different users and items which can enhance the performance and generate accurate and precise recommendations **[14]**. Autoencoder is an unsupervised learning algorithm in which the Artificial Neural Network(ANN) is designed. The input data is reconstructed in the output layer.

In our recommender system, we have employed the models which combine Particle Swarm Optimization(PSO), Fuzzy C-means(FCM) clustering, and the Autoencoder algorithm. PSO is a bio-inspired algorithm that is used for data optimization by reducing dimensionality. The PSO extracts the features and gives the initial cluster positions. FCM is used to classify the users in the dataset using the initial cluster positions. FCM reduces data redundancy. The optimized and clustered data is given as input to the Autoencoder to get the final recommendations of the movies to the users as output. We have proposed two models - one with a single-layer autoencoder and another with a deep autoencoder.

The remaining part of the paper is organized as follows: Section 2 discusses the related work; Section 3 describes the proposed methodology; Section 4 relates to experiments and results; Section 5 relates to the conclusion and future work.

2. Related Work

For the recommendation systems, various solutions use collaborative filtering to help the users to find the items of their interests and also benefit the organizations by improving their customer relations **[14] [15]**. To handle the issues and limitations of recommender systems related to high-dimensionality, data sparsity, complexity, and non-linearity, a recommender system model is used. Clustering and nature-inspired optimization algorithms **[6]** have been used in the past that help in dimensionality reduction and also handle the data sparsity. Data optimization is a challenge for recommender systems, so bio-inspired optimization algorithms work well for this. In the literature, traditional recommender systems are designed based on the matrix factorization and Singular Value Decomposition(SVD)**[16][17]**, clustering techniques, KNN(K-nearest neighbor)**[18]**, nature-inspired optimization algorithms**[19]**, machine learning**[7]**. Recently, various deep learning-based recommender systems have been proposed **[20].** In **[21],** the authors proposed six collaborative filtering-

based item recommendation methods based on different matrix factorization methods and clustering approaches. The six methods are developed by combining a matrix factorization method out of three- Singular Value Decomposition (SVD), Non-Negative Matrix Factorization (NMF), and Principal Component Analysis (PCA), with a clustering approach either K-Means or Fuzzy C-Means(FCM). The model works in three steps- the matrix factorization method to generate the latent factors; the clustering approach to group similar users; and the user-based collaborative filtering item recommendation. The Movielens dataset is used and the results show that the combination of PCA and K-means performs better out of six defined methods. In **[22],** Enhanced Singular Value Decomposition(ESVD), an enhanced version of SVD is proposed with three variants like the Multi-layer ESVD, Item-wise ESVD, and User-wise ESVD. The proposed methodologies are evaluated on Netflix and Movielens data sets and the results are validated in terms of accuracy and efficiency when compared to each other and the existing SVD model. In **[23]**, the authors proposed a collaborative movie-based recommender system with fuzzy C-means clustering(FCM) algorithm and grey wolf optimizer(GWO) on the movielens dataset. The collaborative filtering technique uses FCM and GWO to recommend the movies to the users. The GWO is used for feature reduction by obtaining the initial cluster. The initial cluster is further used by FCM to classify the users based on the rating similarity. This system works well as compared to k-means(KM), Principal Component Analysis(PCA), PCA-k-means, k-means improved, FCM, KM-PSO-FCM(k-means, particle swarm optimizer, fuzzy c-means). The accuracy achieved by these models is lesser than the proposed models. Also, error rates are also higher than proposed models. The proposed models show better accuracy and low error rate by using the machine learning and big data environment. An important feature of privacy can make the system more efficient. In **[24],** the authors proposed a collaborative filtering-based recommender system using the K-means clustering technique and artificial bee colony(ABC) optimization technique on the movielens dataset. The k-means clustering is applied to the dataset for nearest cluster selection and estimating the rating value. ABC is a bio-inspired optimization algorithm that is based upon the intelligent working of honey bees and used for optimization. ABC is used for dimensionality reduction and to reclassify the users. This system works well as compared to the k-means cluster, PCA-GAKM(Genetic algorithm K-means), PCA-k-means, and GAKM Cluster. But this system doesn't consider privacy and cross-domain data context, which can be included in the future work. In **[25]**, an efficient Collaborative Filtering(CF) based recommender system is proposed using Fuzzy C-means clustering(FCM) and Artificial Algae Algorithm(AAA) with an advanced Pearson correlation coefficient(PCC). PCC is used as a similarity measure. First, the FCM was applied to the rating matrix of movies and users. Each user is assigned a cluster. Then PCC was applied to the FCM clusters to get similarities among the users. Then AAA optimization was used to get colonies for each user, similar to bee colonies, from the rating matrix. The intermediate similarities were compared with FCM's similarities. Finally, the similarity of the users is found by using multi-level filters. The items are recommended

according to the similarity of users. This system performed well when compared with k-means (KM), Principal Component Analysis(PCA), PCA-k-means, FCM, KM-PSO-FCM, and PCA-GAKM. The challenge posed to this system for recommendations is to use sentiments, context, and social-based features along with big data. In **[26]**, a collaborative recommender system that is enhanced with particle swarm optimization(PSO) has been proposed. Two clustering algorithms, namely, fuzzy c-means and k-means are used with the PSO algorithm. Also, the type division method is used initially on the dataset to divide the users according to the types of movies they watch. Then k-means and PSO are used to get the precise cluster centers. These cluster centers are provided as input to FCM. This model gives more accurate and reliable results as compared to k-means(KM) clustering, PCA-k-means, PCA-KM, and PCA-GAKM. In **[27]**, a collaborative movie recommender system using clustering and bat optimization has been proposed. This paper combines Fuzzy C-means with Bat optimization. Bat optimization is used to enhance the clustering algorithm and improves the quality of recommendation. Data sparsity and scalability are the main problems that are solved with this model. This model has been compared with k-means, PCA KM, PCA-GAKM, Self-Organizing Map(SOM), UPCC(Pearson Correlation Coefficient based method) and found better results. In **[28]**, a CF-based movie recommender system using K-means and ant colony optimization(ACO) has been proposed. This system performed well and gave accurate recommendations as compared to k-means, PCA KM, PCA-GAKM, Self-Organizing Map(SOM), PCA-SOM, UPCC(Pearson Correlation Coefficient based method), and ABC-KM. The performance was estimated on various parameters such as MAE, accuracy, recall, and precision. To improve the recommendation systems and address the issues in the field of big data and data complexity issues, deep learning techniques have been used by the researchers**[11].** Restricted Boltzmann Machines(RBM) was the very first attempt for recommender systems**[29]**. However, different approaches are being proposed based on Autoencoders **[30]**. In **[31]**, an autoencoder-based collaborative filtering method has been proposed, which is applied to the movielens dataset. For better results and performance, a stacked autoencoder is used with pre-training with RBM. This model focuses on item ratings only. It has shown good performance models such as user-based collaborative filtering, item-based collaborative filtering, Singular Value Decomposition, autoencoder(with and without pre-training), and stacked autoencoder without pretraining. In **[32]**, AutoRec has been proposed which is a collaborative filtering-based recommender system using Autoencoder. The Autorec is an efficient model which outperforms various collaborative filtering techniques like; RBM-CF, biased Matrix Factorization, and Local Low-Rank Matrix Approximation(LLORMA). The datasets used are Movielens and Netflix. In **[33]**, the authors presented a method called Collaborative Denoising Autoencoder(CDAE. It is used for the top n recommendations. This paper is based on the concept of denoising autoencoders which is used to learn the users-items representations. Three datasets have been used and CDAE has been compared with ITEMCF, MF, Bayesian Personalized Ranking(BPR), and Factored Item Similarity Model(FISM). In **[34]**, the authors proposed a hybrid

recommender system using Autoencoders which is a collaborative filtering method and uses side information. Scalability and Cold-start problems are solved using this method. The Datasets used are Movielens and Douban datasets. In **[35]**, the authors proposed a deep learning approach based on autoencoders. This approach is used to predict movie ratings for a user based on other users' ratings. It is a collaborative filtering technique used on the Movielens dataset to recommend the movies. This approach is compared to existing collaborative filtering techniques- K-nearest neighbor(KNN) and Matrix factorization, and it performed well. In **[36]**, a flexible deep autoencoder named FlexEncoder is proposed with new and unique features being embedded to analyze how parameters affect the prediction accuracy of recommender systems. This helps in getting parameters that have the best performance. In **[37]**, a Deep Collaborative Conjunctive recommender(DCCR) system has been proposed. The method is used for the prediction of ratings. A DCCR is a hybrid model of autoencoder and multi-layered perceptron(MLP). Here the autoencoder is used to get the latent features from users-items perspectives in parallel while MLP is used to represent the users-items interaction by fusing the latent features. The sole motive of this model was to combine two different neural networks and improve the accuracy of rating prediction. In **[38]**, a product recommender system using an autoencoder is proposed, which is based on collaborative filtering. This model performed well and resulted in lower RMSE values as compared to Singular Value Decomposition(SVD) based RS. In **[39]**, a deep learning-based algorithm for multi-criteria recommender systems has been proposed. This model uses the Autoencoder-based deep learning approach to make use of non-linear and hidden relationships between the users with their multiple criteria preferences and gives more reliable and accurate recommendations. In **[40]**, the authors proposed a model-Deep Autoencoders for Feature Learning with embeddings for recommendations(DAFERec). This model incorporates the features from autoencoders and combines them with embeddings into a deep neural network. This is done to predict ratings in recommender systems. The limitation of this model is that it requires optimization for the computation of training time and prediction time for this algorithm. In **[41]**, the authors proposed a recommender system which is based on deep sparse autoencoder and particle swarm optimization algorithm. Initially, a deep sparse autoencoder is used to learn latent features based on the ratings matrix, tag information and trust relationships. Then, the particle swarm optimization is applied to find the optimal weights of these latent features. The optimal weights are used further in calculation of unknown ratings. The limitation of this model is that it considers an excess number of latent features which increases the probability of using irrelevant latent features and results in low performance. In **[42]**, the authors proposed a Model-based Collaborative Filtering Algorithm Based on Stacked AutoEncoder (MCFSAE). This model overcomes the sparsity problem. The limitation with this model is that the contextual and content information are not considered. This information is required to construct more reasonable user and item features, which can further alleviate the spasticity issue.

The major contributions of this research work are:

1. The proposed model is a novel recommender system combining three techniques- Fuzzy C-means clustering, Particle Swarm Optimizer, and Autoencoder.

2. Particle Swarm Optimizer is used for feature selection i.e. dimensionality reduction and to get the initial cluster, and Fuzzy C-means algorithm is used to cluster the data and reduce the data redundancy.

3. Deep Learning is used with a recommender system for modeling the latent feature interactions between user and item. Autoencoders are used for the classification.

4. The system's behavior is analyzed using; Mean Squared Error(MSE), Mean Absolute Error (MAE), Precision, Recall and Accuracy.

3. The Proposed Methodology

In this segment, we presented the proposed Recommender System which combines FCM, PSO, and Autoencoder. Two models are designed based on the number of layers of Autoencoder. The proposed model 1 is based on PSO-FCM and a single-layer autoencoder, which is named PFA-RS. Whereas, the proposed model 2 is based on PSO-FCM and Deep layer Autoencoder, which is named PFDA-RS. The movielens dataset is used as an input with users, movies, and their ratings. Both algorithms work in four phases but the difference lies in the second phase. In the first phase, the dataset is prepared and PSO-FCM is applied combinedly. In this process, a data frame is created from the dataset. The dataset is converted to a NumPy array and saved to a CSV text file. The text file is then loaded. The steps of the PSO algorithm are applied which gives the best solution. Now, the cluster centers generated by PSO are considered with FCM as the objective function to be minimized. The users are clustered based on movies watched and given ratings. The cluster number column is added to the data frame. The data frame is split into test, train, and validate data frames. The number of users, ratings, and items are extracted from the data frame. The sparse pivot tables are created with users in rows and movies in columns to test, train, and validate datasets. In the second phase, based on the number of layers both proposed models are designed. For model PFDA-RS, a Deep Autoencoder is constructed. The number of hidden layers, number of neurons in each hidden layer, and number of input and output layers as a number of unique items i.e. movies are defined. The weight of the autoencoder network is initialized. The parameters such as; activation function: sigmoid, loss function: mean square error, learning rate: 0.01, the number of epochs: 500, and weight decay: 0.1 are defined. The optimizer object RMSprop is constructed with

learning rate and weight decay. In the third phase, network training and testing are done. For each epoch and each user and movie, encoding and decoding are done with the calculation of reconstruction error. The autoencoder weights and biases are updated after each observation using the chosen optimizer and loss function. Similarly, for model PFA-RS, a single-layer autoencoder is constructed. After the construction of autoencoders, general criterion-based predictions are given. In the last phase, for each user and all movies, overall rating prediction is computed by taking the average of underlying criterion-based prediction for each movie. Every phase works for both models.

3.1 Particle Swarm Optimization

Kennedy & Eberhart **[43]** proposed a Particle Swarm Optimization (PSO) approach based on swarm behavior observed in nature such as fishes schooling and birds flocking. Particle Swarm Optimizer is an optimization algorithm that is used for optimal decision-making in various domains**[44][45]**. It is a nature-inspired optimization algorithm. Being an evolutionary computation technique, it is based on social behavior. In PSO, the particles fly around in search space which is multidimensional, to find the optimal solution. The population of particles gives random responses. There is a random velocity for every particle by which it is drifted across the n-dimensional problem space. The position and velocity of particles are revised based on the previous velocity, the previous best location of the particle and the previous best global location of the particle. The main objective of PSO is to accelerate each particle towards the best location and global location of the particle. The velocity and position of the particle are given equations 1 and 2 respectively:

$$v_{in} = w \times v_{in} + c_1 \times rand_1() \times (p_{in} - x_{in}) + c_2 \times rand_2() \times (p_{gn} - x_{in}) \quad (1)$$

$$x_{in} = x_{in} + v_{in} \quad (2)$$

$where\ l_{in} = the\ position\ of\ each\ particle,$

$v_{in} = the\ velocity\ for\ each\ particle,$

$i = the\ particle,$

$n = the\ dimension,$

$p_{in} = the\ previous\ best\ location\ of\ the\ particle,$

$$p_{gn} = the\ previous\ best\ global\ location\ of\ the\ particle$$

3.2 Fuzzy C-means

Fuzzy C-means is a clustering algorithm based on fuzzy logic. FCM allows the items to be in more than one cluster with a membership percentage value. In our proposed system, we have used FCM to cluster the data**[46]**. In this algorithm, the minimization of the following objective function is done in equation 3:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{V} \mu_{ij}^{m} \|x_i - v_j\|^2 \quad (3)$$

$$where\ m = real\ number\ greater\ than\ 1,$$

$$N = number\ of\ items,$$

$$V = the\ number\ of\ clusters,$$

$$\mu_{ij} = degree\ of\ membership\ of\ an\ item,$$

$$i = item,$$

$$j = cluster,$$

$$x_i = the\ i^{th} value\ of\ d - dimensional\ data,$$

$$v_j = the\ cluster's\ d - dimensional\ center$$

The cluster Center V_{ij} is calculated in equation 4 :

$$V_j = \frac{\sum_{i=1}^{N} \mu_{ij}^{m} . x_i}{\sum_{i=1}^{N} \mu_{ij}} \quad (4)$$

The membership μ_{ij} is calculated in equation 5 :

$$\mu_{ij} = \sum_{k=1}^{V} \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|}\right)^{\frac{-2}{m-1}} \quad (5)$$

k is the number of iterations which stops when a condition $max_{ij}\{|\mu_{ij}(k+1) - \mu_{ij}k|\} < \epsilon$ *is met, where* ϵ *is a criteria to terminate and value is between 0 and 1. This process continues until it converges to a local minima point or a point* J_m **[47]**.

3.3 Autoencoder

Autoencoder is a peculiar type of feed-forward neural network which has the same input and output. There are 3 components- encoder, code and decoder. The input from the encoder is converted to a low-dimensional code which is further used by the decoder to reconstruct the input. A loss function is used to compare the output with the input. Autoencoder is mainly used for dimensionality reduction and information retrieval. Its important features are- data specific, lossy and unsupervised. Both encoder and decoder are the Artificial Neural Networks(ANN'S) i.e. fully-connected networks which are feed-forwarded. The code layer has a single layer of ANN. The number of code layers is always less than the encoder and decoder layers. The code layer has a dimensionality which we choose based on setting the hyperparameter before training the autoencoder **[38]**. The structure of the autoencoder is shown in Fig 1.

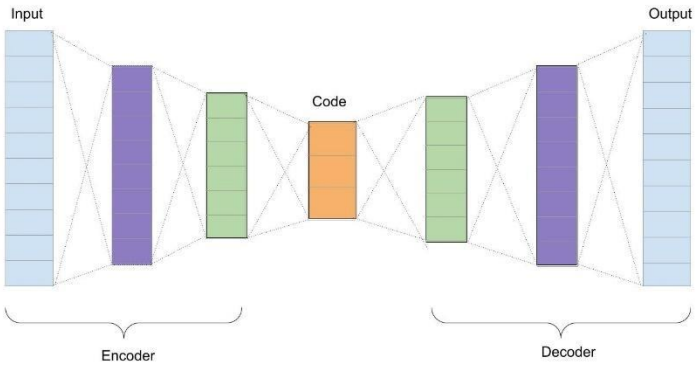


Fig.1 Autoencoder

The autoencoder can be as deep as we may want it. In Fig 1., we have two layers in both encoder and decoder without considering the input and output, a single layer for code or we can call it a hidden layer. Here, let i represents the input data, c represents the code layer data or hidden layer data and i' represents the output data. There are two main parts of autoencoder- one is encoder which maps the input to the code and other is decoder which maps the code to reconstruct the actual input as the output as shown in equations 6 &7 respectively.

$$h = g(Wx + b) \tag{6}$$

$$x' = f(W_1\, c + b_1) \tag{7}$$

$where\ g(.)\ and\ f(.) = the\ non - linear\ activation\ functions,$

$x = the\ input\ data,$

$x' = the\ output\ data,$

$h = the\ encoded\ data,$

$W = the\ weight\ matrix\ between\ input\ layer\ and\ the\ code\ layer,$

$W_1 = the\ weight\ matrix\ between\ code\ layer\ and\ output\ layer,$

$b\ and\ b_1 = the\ bias\ vectors\ of\ hidden(or\ code\ layer) and\ the\ output\ layer$ [37]

Autoencoder minimizes the reconstruction error between i and i' for which we use L1 loss function which is as follows in equation 8 **[37]:**

$$L(x, x') = \|x - x'\| = \|x - f(W_x + b) - b_1\| \tag{8}$$

In our work, we have used single layer and deep layer autoencoders. Single layer autoencoder is represented in Fig.2 , and deep layer autoencoder is represented in Fig.3.

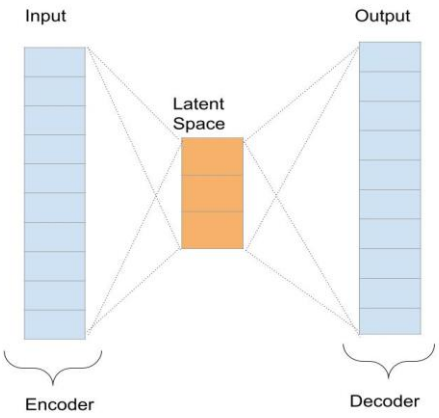


Fig. 2 Single layer based autoencoder

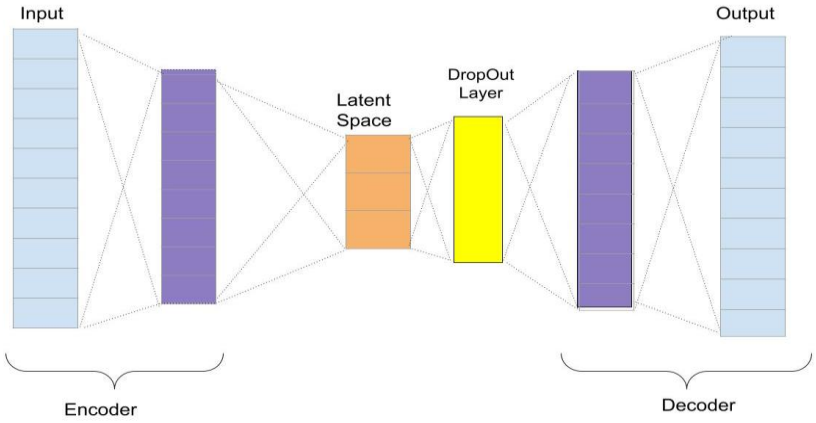


Fig. 3 Deep layer based autoencoder

The workflow of the system for single layer Autoencoder based model is shown in Fig. 4, and for deep layer Autoencoder based model is shown in Fig.5. Initially the Movielens dataset is loaded and the data is processed by removing the duplicate values and transforming the data through normalization. Then, this processed data is used as an input to the particle swarm optimization algorithm to optimize the data. Using the particle swarm optimization, the dimensionality reduction is performed. This optimized data is used as an input to the fuzzy c-means clustering algorithm to get the clusters of data. The clustering of data helps in using similar kinds of data. The clustered data is used as an input to a deep learning model called Autoencoder. An Autoencoder is a neural network that is trained to copy the input to the output for better optimization and

reduced dimensionality of the data. The difference in both the models lies in the layers of autoencoder. In the first model, only a single input-output layer pair with a hidden layer is used, while in the second model, there are many deep layers i.e. input-output layer pairs with more hidden layers.

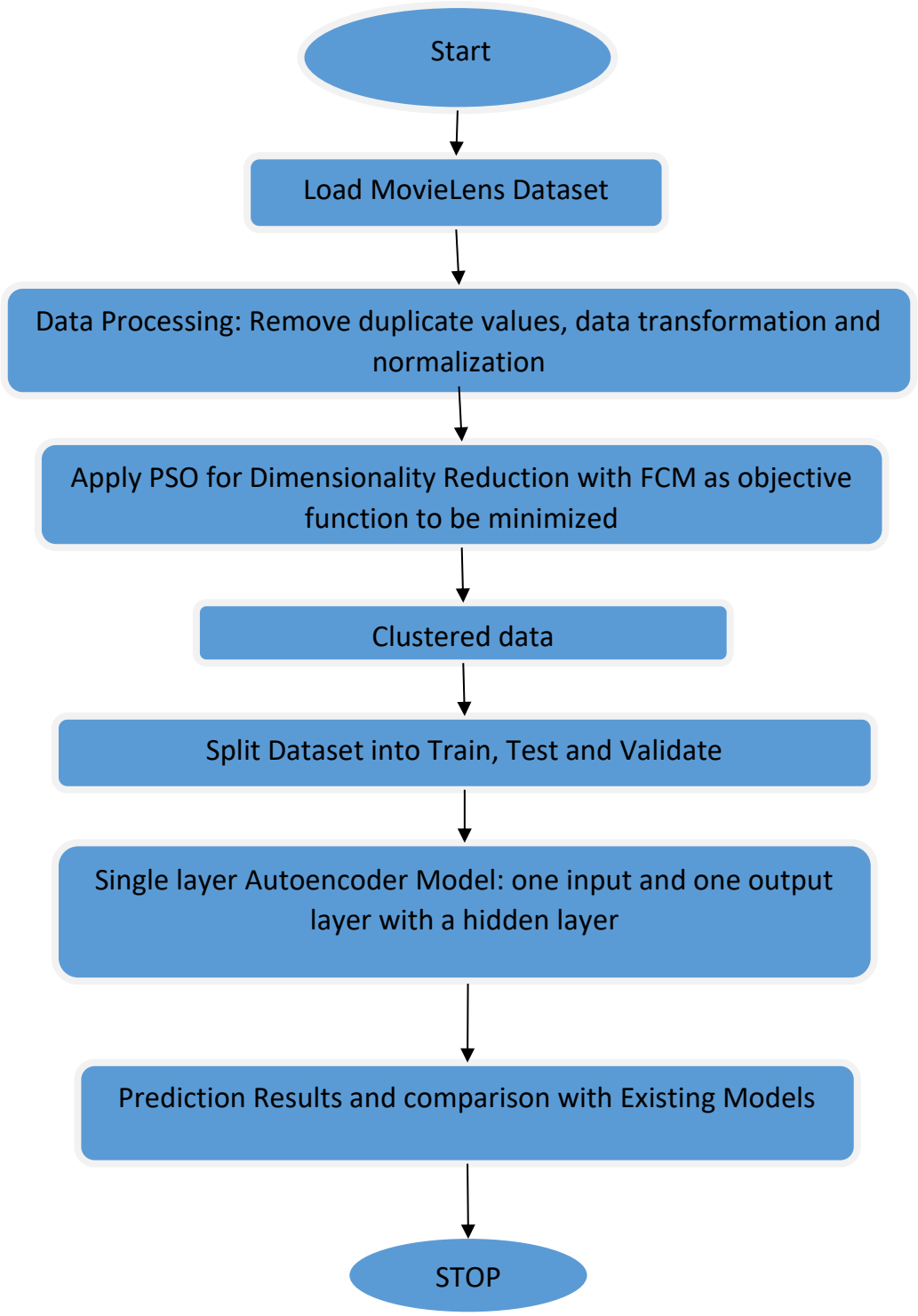


Fig.4. Flowchart of single layer autoencoder based movie recommender system

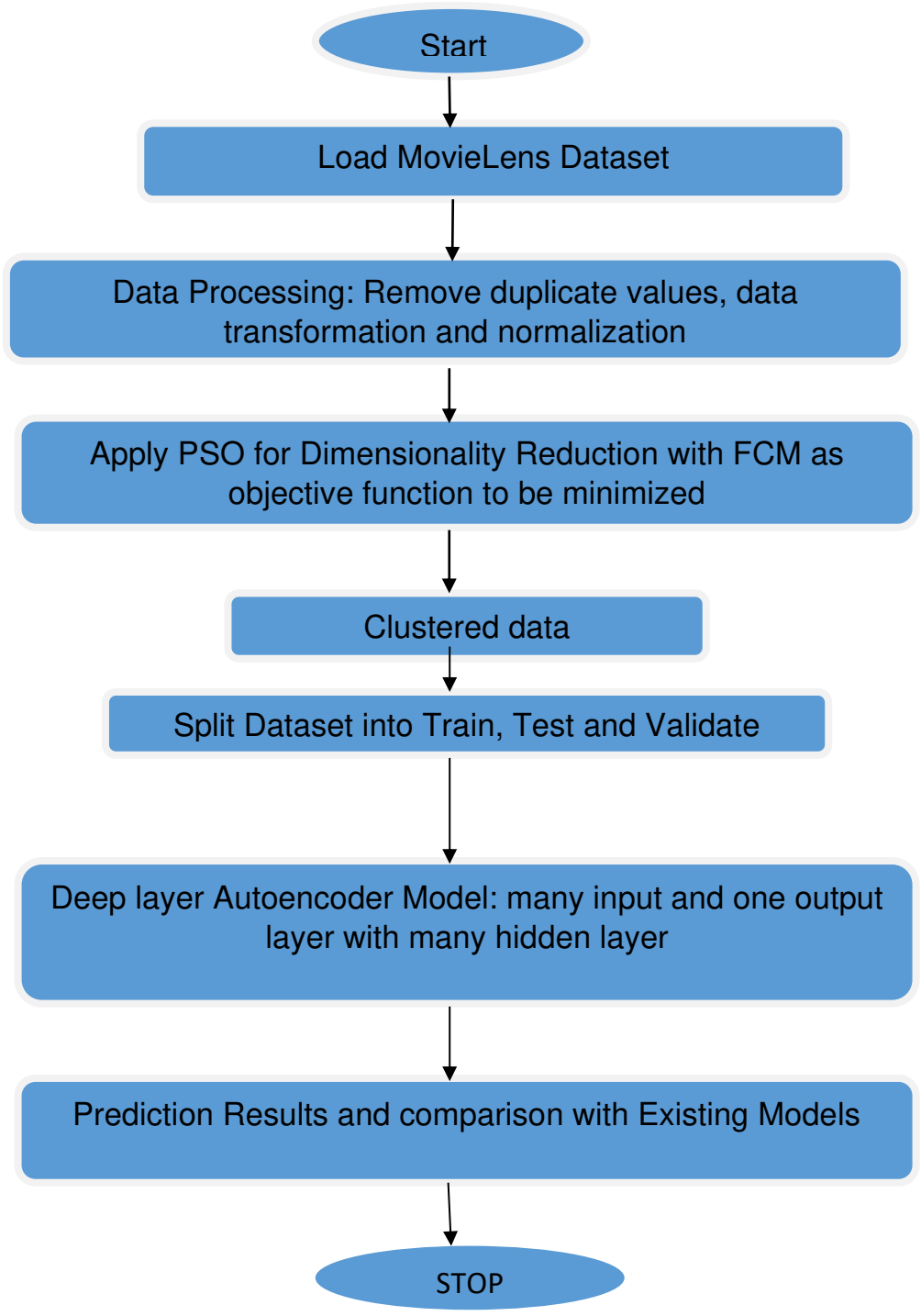


Fig.5. Flowchart of deep layer autoencoder based movie recommender system

The algorithm of our proposed system is described in Fig.6.

Input: Dataset D(u*m*r) // u: user, m:movie, r:rating

Output: P(u*m) // overall predictions of user-movie ratings

Phase 1: Dataset Preparation, Particle Swarm Optimizer and Fuzzy C-means Clustering

1: Create a dataframe DF(u*m*r) from the dataset D(u*m*r).

2: Convert the dataset to numpy array and Save the numpy array to a csv text file

3: Load the text file

4: For every user i = 1, ...,S do:

4a: Initialize the user's position with a vector(random): ki ~ (bl, bu), where bl and bu are the lower and upper bound of the problem search-space.

4b: Initialize the user's position to its initial position: pi ← ki

5: If (f(pi) < f(g)) refresh the swarm's finest known position: g ←pi

5a: Initialize the user's property(velocity): vi ~ U(|bu-bl|, |bu-bl|)

6: Do while terminating criterion is not reached:

6a: For every user i = 1, ...,S do:

6aa: take random numbers: rp, rg ~ U(0,1)

6b: For every dimension d = 1, ..., n do:

6ba: Update the user's property(velocity): v(i,d) ← ω v(i,d) + ϕp rp (p(i,d)- k(i,d)) + ϕg rg (gd- k(i,d))

6bb: Update the user's position: ki ← ki + vi

7: If (f(xi) < f(pi)) do:

7a: Update the user's best-known position: pi ← ki

8: If (f(pi) < f(g))

8a: Update the swarm's greatest known position: g ← pi

9: g holds the best-found solution.

10: Take cluster centers generated by PSO with FCM as objective function to be minimised

11: The users are clustered based on movie watched and ratings given.

12: Cluster number column is added to the dataframe.

13: Split the dataframe into train, test and validate dataframes.

14: Extracting the number of ratings,users and items from D(u*m*r)

15: Creation of sparse pivot tables with users in rows and movies in columns for train,test and validate datasets and preprocessing the data.

Phase 2: Construction of Deep AutoEncoder

1: Define the parameters: number of hidden layers ; neurons in each hidden layer; number of input and output layers as number of unique items i.e movies.

2: Initialize the weight of the AutoEncoder network.

3: Define the activation function: sigmoid; loss function: Mean square Error; learning rate: 0.01; number of epochs: 500; weight decay: 0.1

4: Construct the Optimizer object(RMSprop) with the learning rate, weight decay

Phase 3: Network Training and Testing

1: For each Epoch do

1a: For each User for all movies do

1aa: Encode the user rating input into another vector V having low dimension than the input(eqn 1)

1ab: Decode the vector V to an output vector having same dimension as the input(eqn 2)

1ac: Calculate the reconstruction error(equation 3)

1ad: Update the autoencoder weights and biases after each observation(Stochastic Gradient Descent) using the chosen optimizer and loss function

2: end For

3: General Criterion based predictions

Phase 4: Overall Predictions

1: For each User for all movies do

1a: Compute overall rating prediction by taking the average of underlying criterion based prediction for each movie

2: end For

Fig.6. Algorithm of proposed movie recommender system

The PSO, FCM and Autoencoder have been used in combination as such one's result is fed to another as input. Initially, the PSO is applied to the movielens dataset and results in the optimized dataset with most relevant attributes. The optimized dataset is used as input to the FCM to cluster the data based on their similar interests. The similar kind of data will be in the same cluster. The clustered data is well organized data which is now used as input to the autoencoder to make our data more optimized. Since autoencoders are used for information retrieval and perform prediction. So, an autoencoder is applied to the clustered and optimized data, to get the recommendations of movies to the users based on their interests. We have proposed two versions of our model- one with single layer autoencoder and other with deep layers of autoencoder. The two models are named as- PSO-FCM and Autoencoder based Recommender System(PFA-RS) and PSO-FCM and Deep Autoencoder based Recommender System(PFDA-RS). Various parameters are used for the Autoencoder model. These parameters are: Activation Function- Selu/Linear, Optimizer- Adaptive Moment Estimation Optimizer(ADAM), Loss Function- Mean Absolute Error(MAE), metrics- Accuracy, Mean Square Error(MSE), Precision, Recall and Dropout.

4. Dataset and Evaluation Measures

The Movielens dataset is used to evaluate our proposed recommender system, which is freely available. The dataset was collected by The University of Minnesota for the Grouplens Research Project. The dataset contains 100K ratings from 1 to 5 and 943 users. The users have rated 1682 movies. Some evaluation metrics have been considered for better understanding the quality of the recommendations by the user. We calculated the mean square error(MSE), mean absolute error (MAE), precision, recall and accuracy to measure the performance of the proposed system. If the value of MAE is small, then prediction accuracy is good. So, the MAE should be less to get high accuracy value for a recommender system.

$$MAE = \frac{\sum \|p_{ui} - r_{ui}\|}{M} \quad (9)$$

$$where\ M = Total\ Expected\ Movies,\ \ p_{ui} = Predicting\ rating\ for\ user\ u\ on\ item\ i,\ \ r_{ui} = Real\ rating$$

$$MSE = \frac{1}{M}\sum_{i=1}^{M}(p_{ui} - r_{ui})^2 \quad (10)$$

$$Precision = \frac{|Interesting \cap TopN|}{N} \quad (11)$$

$$Recall = \frac{|Interesting \cap TopN|}{|Interesting|} \quad (12)$$

Precision and recall are used together to evaluate the performance of information retrieval systems and classification models. Precision is defined as the fraction of important instances among all the retrieved instances. Recall, also known as the 'sensitivity', is the fraction of important instances that were retrieved.

In the recommendation system, accuracy gives the computation results about how accurate the recommendations are. The accuracy in terms of True Positive (TP), False Positive(FP), True Negative(TN) and False Negative(FN) is:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

5. Results and Comparison Analysis

In this segment, we will discuss the outcomes and comparisons which were performed among various existing systems by using Movielens dataset. The proposed models are PFA-RS and PFDA-RS. Initially, the datapoints are plotted based on ratings given by users to the movies. The userId is on x-axis which depicts the users in the dataset. The movieId is on y-axis which depicts the movies in the dataset. The rating for a particular movie and user are shown as coloured data points in Fig. 1. The red colour is for rating 0.0, orange colour is for rating 1.5, grey colour is for rating 3.0, the light blue colour is for rating 4.5 and dark blue colour is for rating 6.0. PSO and FCM are applied over the dataset which results into optimized clusters of data. The data points are plotted based on userId and movieId for the ratings given by a particular user for a movie. The data points before clustering are shown in Fig. 7 and the data points after clustering are shown in Fig.8.

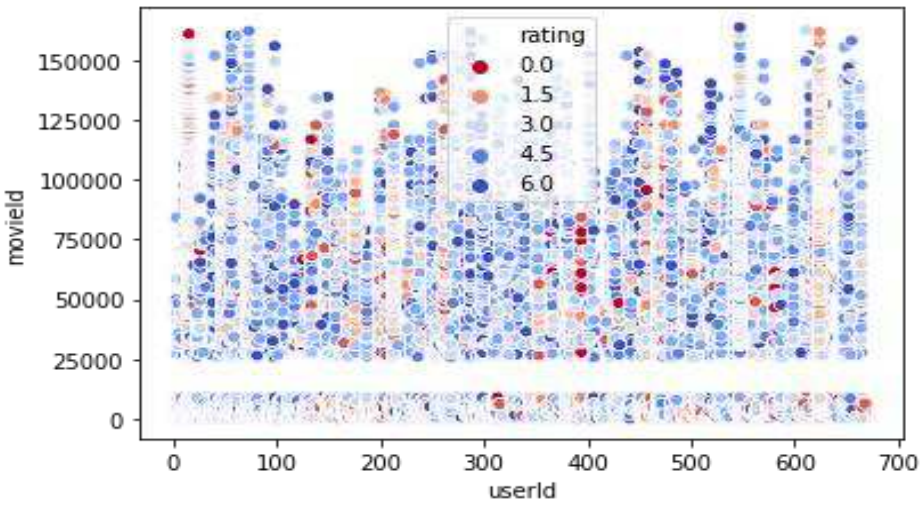


Fig. 7 Plotting of ratings given by users for Movies

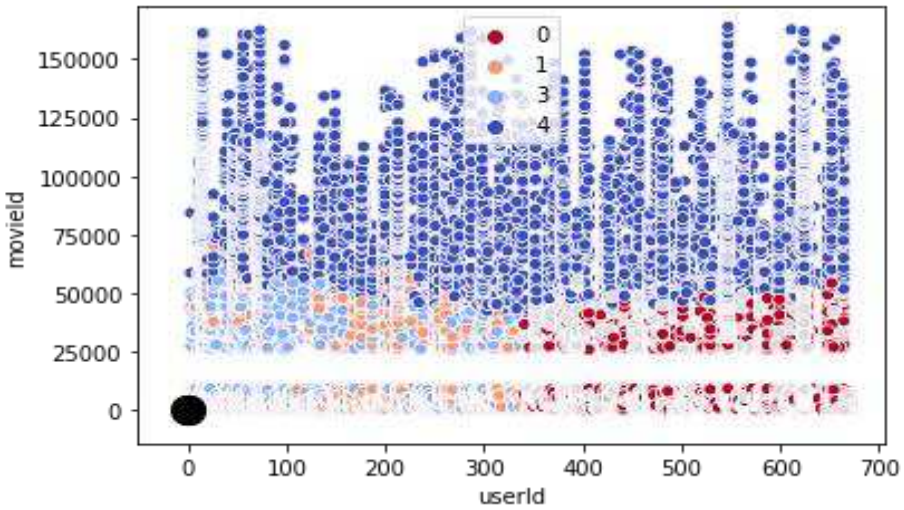


Fig.8 Plotting of clusters of ratings of movies given by users

The summary of the single layer based Autoencoder model is summarized in table 1 and the summary of the deep layer based Autoencoder model is summarized in table 2. The type of layer is mentioned with the output shape of the particular layer. For each layer the number of parameters are calculated. The total number of parameters is the sum of all the weights and biases on the neural network. The number of layers used in Single Layer Autoencoder is one for both encoder and decoder; and a latent or hidden layer is placed between encoder and decoder. As we increase the number of layers in encoder and decoder, it is called deep autoencoder. In Deep Autoencoder, there are 2 layers in both encoder and decoder with a latent layer and a dropout layer in the middle.

Table 1 Summary of single layer based autoencoder model

LAYER(TYPE)	**OUTPUT SHAPE**	**NUMBER OF PARAMETERS**
UserRating (InputLayer)	[(None, 163950)]	0
LatentSpace (Dense)	(None, 500)	81975500
UserScorePred (Dense)	(None, 163950)	82138950
Total params: 164,114,450 Trainable params: 164,114,450 Non-trainable params: 0		

Table 2 Summary of deep layer based autoencoder model

LAYER(TYPE)	**OUTPUT SHAPE**	**NUMBER OF PARAMETERS**
UserRating (InputLayer)	[(None, 163950)]	0
EncLayer0 (Dense)	(None, 256)	41971456
LatentSpace (Dense)	(None, 512)	131584
dropout_4 (Dropout)	(None, 512)	0
DecLayer0 (Dense)	(None, 256)	131328
UserScorePred (Dense)	(None, 163950)	42135150
Total params: 84,369,518 Trainable params: 84,369,518 Non-trainable params: 0		

The model train and validation accuracy of proposed models and other two existing models- AE-RS(Autoencoder recommender system) and DA-RS(Deep autoencoder recommender system), are shown in Fig.9- Fig.12. The change in train and validation accuracy of the proposed and other two existing models are shown with the epochs. As the number of epochs increases, the accuracy value changes. These changes are shown in the line graphs.

Fig.9 shows accuracy vs epochs line graph for a PSO-FCM and Deep Autoencoder based Recommender System. For this model, the train and validation accuracy are compared whose values initially show changes but after a certain epoch the value becomes stable. The train accuracy is better than the validation accuracy.

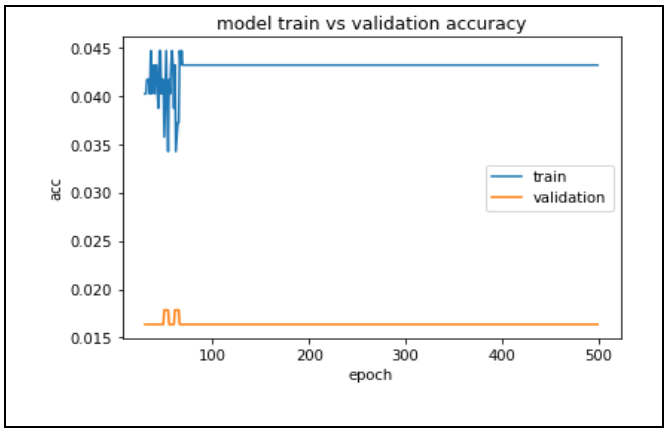


Fig.9 PSO-FCM and deep autoencoder based recommender system (PFDA-RS)

Fig.10 shows accuracy vs epochs line graph for a PSO-FCM and Autoencoder based Recommender System. For this model, the train and validation accuracy are compared whose values initially are stable for certain epochs but show great changes after that. At last, the train accuracy is better than the validation accuracy.

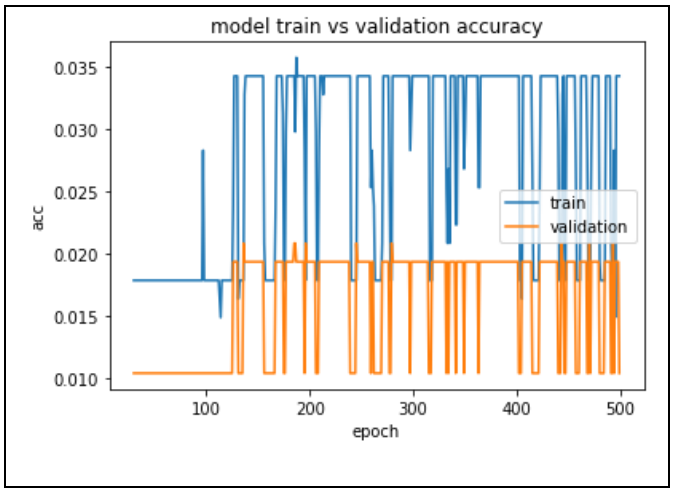


Fig.10 PSO-FCM and single autoencoder based recommender system (PFA-RS)

Fig.11 shows accuracy vs epochs line graph for a Single Layer Autoencoder based Recommender System. For this model, the train and validation accuracy are compared whose values keep changing mostly for every epoch. There are only some values which show stability. The train accuracy is better than the validation accuracy.

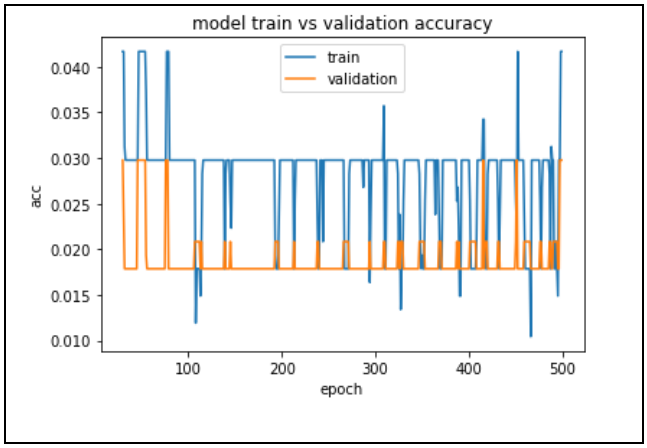


Fig.11 Single layer autoencoder based recommender system(AE-RS)

Fig.12 shows accuracy vs epochs line graph for a Deep Autoencoder based Recommender System. For this model, the train and validation accuracy are compared whose train accuracy values change initially for some epochs and become stable in between. Then, there is variable change in accuracy values after stability. The train accuracy is better than the validation accuracy.

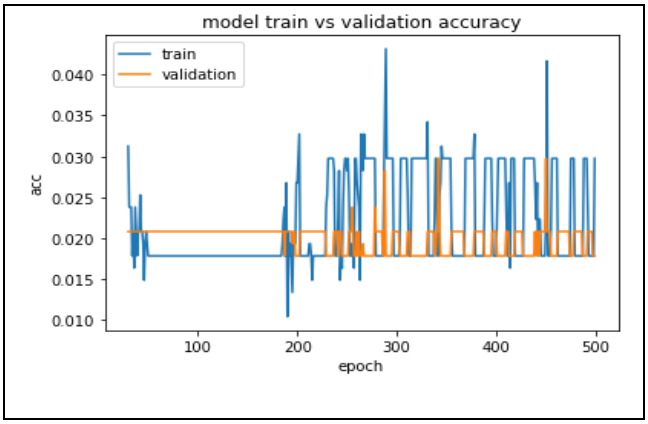


Fig.12 Deep autoencoder based recommender system(DA-RS)

The proposed models-PFDA-RS AND PFA-RS are compared with six existing models- AE-RS, DA-RS, PF-RS, DAFERec, MCFSAE, and DSAPSO. The comparison of results are shown in Fig.13 - Fig.17. In Fig.13, the training mean squared error(MSE) of proposed and existing models is compared in the bar plot. The MSE values of PFDA-RS, PFA-RS, AE-RS **[38]**, DA-RS **[38][39]**, PF-RS, DAFERec **[40]**, MCFSAE **[42]**, and DSAPSO **[41]** are 0.0086, 0.0085, 0.0099,

0.0098, 0.0099, 0.0085, 0.01123, 0.0089 respectively. The MSE of both of the proposed models is lower than the other six existing models.

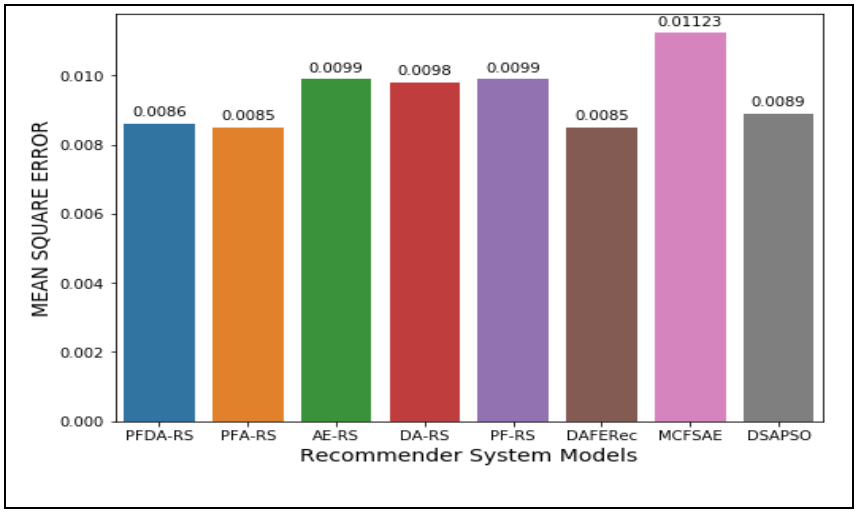


Fig.13 Comparison of training mean square error of recommender system models

In Fig.14, the mean absolute error(MAE) of proposed and existing models is compared in the bar plot. The MAE values of PFDA-RS, PFA-RS, AE-RS, DA-RS, PF-RS, DAFERec, MCFSAE, and DSAPSO are 0.0035, 0.0037, 0.0044, 0.0039, 0.0042, 0.0066, 0.0078, 0.0045 respectively. The MAE of both of the proposed models is lower than the other six existing models.

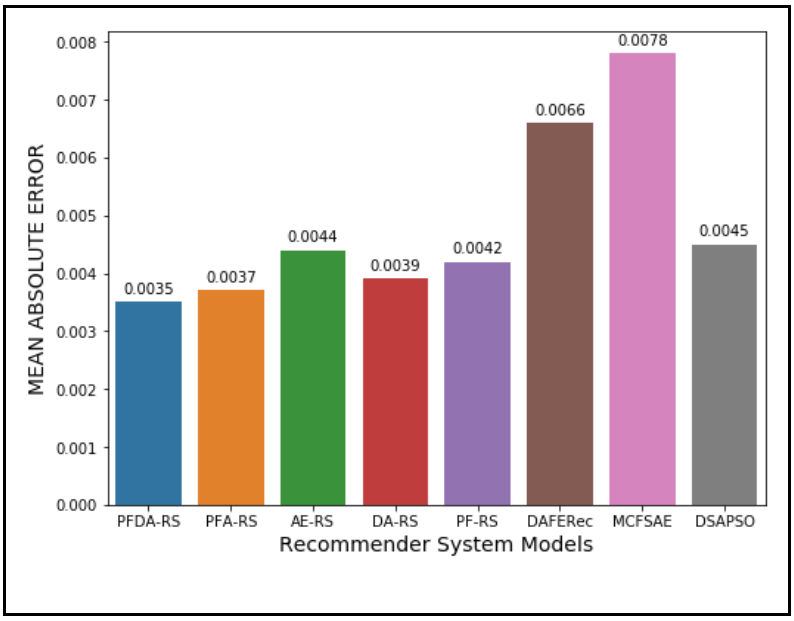


Fig.14 Comparison of mean absolute error of recommender system models

In Fig.15, the precision of proposed and existing models is compared in the bar plot. The Precision values of PFDA-RS, PFA-RS, AE-RS, DA-RS, PF-RS, DAFERec, MCFSAE, and DSAPSO are 0.2278, 0.2164, 0.167, 0.201, 0.12, 0.21, 0.156, 0.199 respectively. The precision of both of the proposed models is higher than the other six existing models.

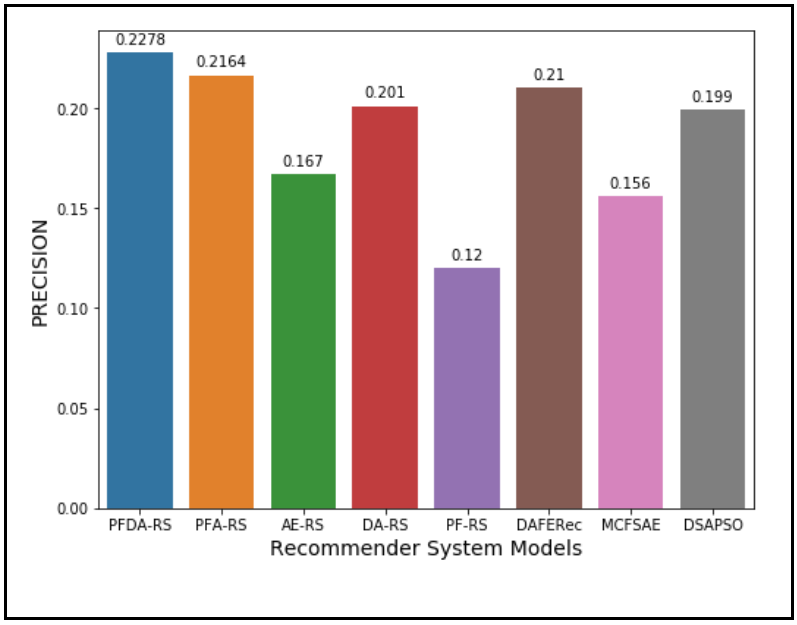


Fig.15 Comparison of precision of recommender system models

In Fig.16, the recall of proposed and existing models is compared in the bar plot. The Recall values of PFDA-RS, PFA-RS, AE-RS, DA-RS, PF-RS, DAFERec, MCFSAE, and DSAPSO are 0.2267, 0.1449, 0.1461, 0.1643, 0.1552, 0.1389, 0.1356, 0.124 respectively. The recall values for proposed model- PFDARS is high other models. The recall value should be between 0.0 and 1.0. The higher the recall value, the model will be able to detect more true positive samples.

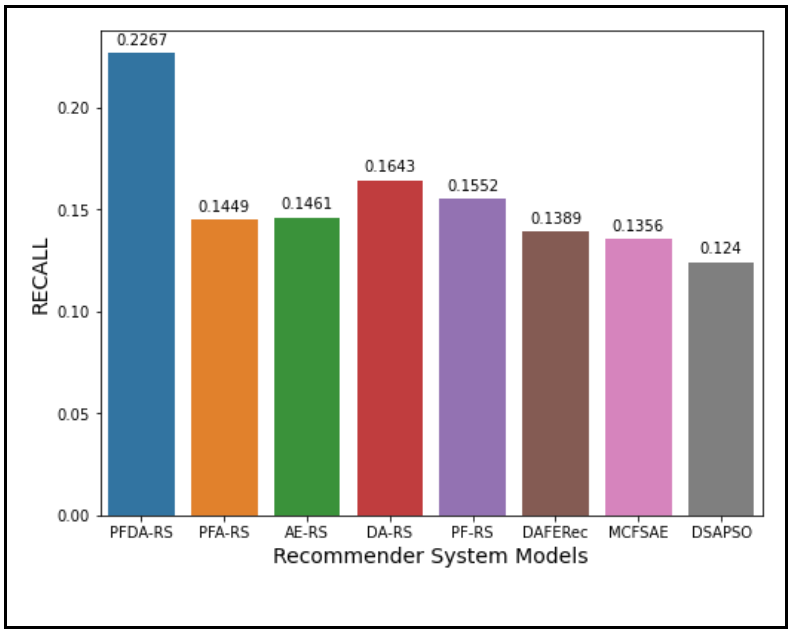


Fig.16 Comparison of recall of recommender system models

In Fig.17, the training accuracy of proposed and existing models is compared in the bar plot. The Accuracy values of PFDA-RS, PFA-RS, AE-RS, DA-RS, PF-RS, DAFERec, MCFSAE, and DSAPSO are 0.0536, 0.0506, 0.0134, 0.0298,

0.0138, 0.0425, 0.0147, 0.034 respectively. The accuracy of both of the proposed models is higher than the other six existing models.

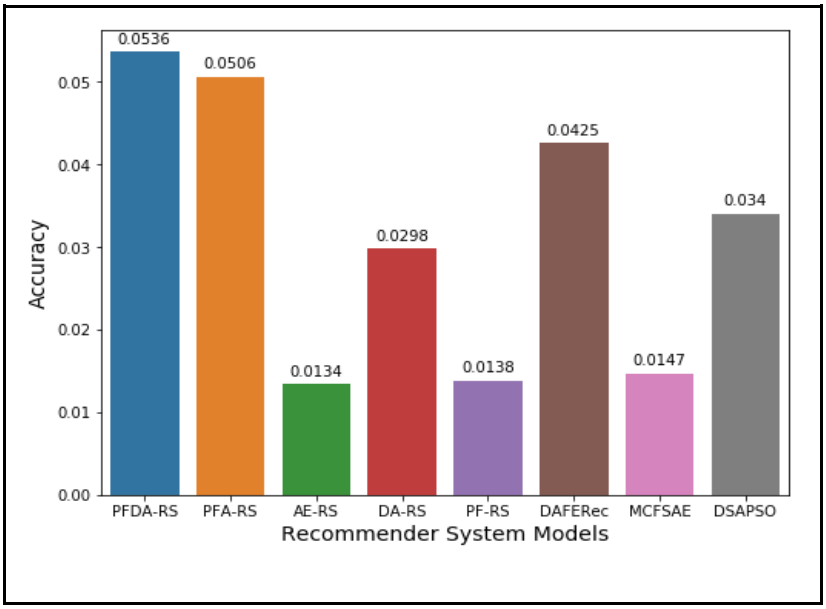


Fig.17 Comparison of training accuracy of recommender system models

The proposed models and existing models have been compared for the values of mean absolute error, mean squared error, precision, recall and accuracy. Both of the proposed models- PFDA-RS and PFA-RS performed well over existing models with best accuracy and low errors for the Movielens dataset. When both models are compared with each other, PFDA-RS works very well.

6. Conclusion

Deep learning has impacted several areas of machine learning. So, it is impacting recommender systems too. In this paper, two recommender systems namely; Particle Swarm Optimizer, Fuzzy C- means Clustering and Deep Autoencoder based Recommender System (PFDA-RS) and Particle Swarm Optimizer, Fuzzy C- means Clustering and Single layer Autoencoder based Recommender System (PFA-RS), have been proposed. Particle Swarm Optimizer is used for the feature selection i.e. dimensionality reduction; and to get the initial cluster. The Fuzzy C-means Clustering algorithm is used to cluster the data and reduce the data redundancy. Deep Learning is used with a recommender system for modeling the latent feature interactions between user and item. Autoencoders are used for the classification. Both models performed extremely well as compared to the existing models like Autoencoder based Recommender System (AE-RS), Deep Autoencoder based Recommender System (DA-RS), Particle Swarm Optimizer and Fuzzy C- means Clustering based Recommender System (PF-RS), Deep Autoencoders for Feature Learning with embeddings for recommendations (DAFERec), Deep sparse

autoencoder and particle swarm optimization based recommender system (DSAPSO), and Model-based Collaborative Filtering Algorithm Based on Stacked AutoEncoder (MCFSAE). The models have been compared on parameters such as Mean Squared Error, Mean Absolute Error, Precision, Recall, and Accuracy. The dataset used is Movielens dataset. For future work, the deep learning models can be applied on different datasets other than the text dataset. The recall and precision values need to be improved for the model.

7. References

[1] Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-based systems*, *46*, 109-132. doi:10.1016/j.knosys.2013.03.012

[2] Rahul Katarya & Om Prakash Verma. (2017). An effective web page recommender system with fuzzy c-mean clustering. Multimedia Tools Appl. 76, 20, 21481–21496. https://doi.org/10.1007/s11042-016-4078-7

[3] Rahul Katarya & Om Prakash Verma. (2016).Recent developments in affective recommender systems. Physica A: Statistical Mechanics its Applications, 461:182–90.

[4] Arpan Kumar Kar, Bio inspired computing – A review of algorithms and scope of applications, Expert Systems with Applications, Volume 59, 2016, Pages 20-32, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2016.04.018.

[5] Rahul Katarya & Om Prakash Verma.(2017) An effective collaborative movie recommender system with cuckoo search, Egyptian Informatics Journal, 105-112.

[6] Parsopoulos, K. E., & Vrahatis, M. N. (2008). Multi-Objective Particles Swarm Optimization Approaches. In L. Thu Bui, & S. Alam (Ed.), *Multi-Objective Optimization in Computational Intelligence: Theory and Practice* (pp. 20-42). IGI Global. http://doi:10.4018/978-1-59904-498-9.ch002

[7] Mu, R. (2018). A survey of recommender systems based on deep learning. *Ieee Access*, *6*, 69009-69022.

[8] Batmaz, Z., Yurekli, A., Bilge, A., & Kaleli, C. (2019). A review on deep learning for recommender systems: challenges and remedies. *Artificial Intelligence Review*, *52*(1), 1-37.

[9] Ganguli, S., & Thakur, S. (2020, January). Machine Learning Based Recommendation System. In *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 660-664). IEEE.

[10] Fakhfakh, R., Ammar, A. B., & Amar, C. B. (2017). Deep learning-based recommendation: Current issues and challenges. *International Journal of Advanced Computer Science and Applications (IJACSA)*, *8*, 12.

[11] Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys (CSUR)*, *52*(1), 1-38.

[12] Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and trends in signal processing*, *7*(3–4), 197-387.

[13] Kiran, R., Kumar, P., & Bhasker, B. (2020). DNNRec: A novel deep learning based hybrid recommender system. *Expert Systems with Applications*, *144*, 113054.

[14] He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017, April). Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web* (pp. 173-182).

[15] Wang, H., Wang, N., & Yeung, D. Y. (2015, August). Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1235-1244).

[16] Y. Koren, R. Bell and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," in *Computer*, vol. 42, no. 8, pp. 30-37, Aug. 2009, doi: 10.1109/MC.2009.263.

[17] R. Mehta and K. Rana, "A review on matrix factorization techniques in recommender systems," *2017 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA)*, 2017, pp. 269-274, doi: 10.1109/CSCITA.2017.8066567.

[18] W. Liang, G. Lu, X. Ji, J. Li and D. Yuan, "Difference factor' KNN collaborative filtering recommendation algorithm", *International Conference on Advanced Data Mining and Applications*, pp. 175-184, 2014, December.

[19] Darwish, A. (2018). Bio-inspired computing: Algorithms review, deep analysis, and the scope of applications. *Future Computing and Informatics Journal*, *3*(2), 231-246.

[20] Betru, B. T., Onana, C. A., & Batchakui, B. (2017). Deep learning methods on recommender system: a survey of state-of-the-art. *International Journal of Computer Applications*, *162*(10), 17-22.

[21] N. Ifada, M. K. Sophan, M. N. Fitriantama and S. Wahyuni, "Collaborative Filtering Item Recommendation Methods based on Matrix Factorization and Clustering Approaches," *2020 10th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)*, 2020, pp. 226-230, doi: 10.1109/EECCIS49483.2020.9263450.

[22] X. Guan, C. -T. Li and Y. Guan, "Matrix Factorization With Rating Completion: An Enhanced SVD Model for Collaborative Filtering Recommender Systems," in *IEEE Access*, vol. 5, pp. 27668-27678, 2017, doi: 10.1109/ACCESS.2017.2772226.

[23] Katarya, R., Verma, O.P. Recommender system with grey wolf optimizer and FCM. Neural Comput & Applic 30, 1679–1687 (2018). https://doi.org/10.1007/s00521-016-2817-3

[24] Katarya, R. Movie recommender system with metaheuristic artificial bee. Neural Comput & Applic 30, 1983–1990 (2018). https://doi.org/10.1007/s00521-017-3338-4

[25] Rahul Katarya, Om Prakash Verma, Effectual recommendations using artificial algae algorithm and fuzzy c-mean, Swarm and Evolutionary Computation, Volume 36, 2017, Pages 52-61, ISSN 2210-6502, https://doi.org/10.1016/j.swevo.2017.04.004.

[26] Katarya, R., & Verma, O. P. (2016). A collaborative recommender system enhanced with particle swarm optimization technique. *Multimedia Tools and Applications*, *75*(15), 9225-9239.

[27] Vellaichamy, V., & Kalimuthu, V. (2017). Hybrid Collaborative Movie Recommender System Using Clustering and Bat Optimization. International Journal of Intelligent Engineering and Systems, 10(5), 38–47.

[28] Kumar, M. S., & Prabhu, J. (2020). A hybrid model collaborative movie recommendation system using K-means clustering with ant colony optimisation. *International Journal of Internet Technology and Secured Transactions*, *10*(3), 337-354.

[29] Salakhutdinov, R., Mnih, A., & Hinton, G. (2007, June). Restricted Boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning* (pp. 791-798).

[30] Zhang, G., Liu, Y., & Jin, X. (2020). A survey of autoencoder-based recommender systems. *Frontiers of Computer Science*, *14*(2), 430-450.

[31] Ouyang, Y., Liu, W., Rong, W., & Xiong, Z. (2014, November). Autoencoder-based collaborative filtering. In *International Conference on Neural Information Processing* (pp. 284-291). Springer, Cham.

[32] Sedhain, S., Menon, A. K., Sanner, S., & Xie, L. (2015, May). Autorec: Autoencoders meet collaborative filtering. In *Proceedings of the 24th international conference on World Wide Web* (pp. 111-112).

[33] Wu, Y., DuBois, C., Zheng, A. X., & Ester, M. (2016, February). Collaborative denoising auto-encoders for top-n recommender systems. In *Proceedings of the ninth ACM international conference on web search and data mining* (pp. 153-162).

[34] Florian Strub, Romaric Gaudel, Jérémie Mary. Hybrid Recommender System based on Autoencoders, the 1st Workshop on Deep Learning for Recommender Systems, Sep 2016, Boston, United States, pp.11 - 16, 10.1145/2988450.2988456. hal-01336912v2

[35] Lund, J., & Ng, Y. K. (2018, July). Movie recommendations using the deep learning approach. In *2018 IEEE International Conference on Information Reuse and Integration (IRI)* (pp. 47-54). IEEE.

[36] Dai Tran, H., Hussain, Z., Zhang, W. E., Dang Khoa, N. L., Tran, N. H., & Sheng, Q. Z. (2018). Deep Autoencoder for Recommender Systems: Parameter Influence Analysis. *arXiv e-prints*, arXiv-1901.

[37] Wang, Q., Peng, B., Shi, X., Shang, T., & Shang, M. (2019). DCCR: deep collaborative conjunctive recommender for rating prediction. *IEEE Access*, *7*, 60186-60198.

[38] Ferreira, D., Silva, S., Abelha, A., & Machado, J. (2020). Recommendation system using autoencoders. *Applied Sciences*, *10*(16), 5510.

[39] Shambour, Q. (2021). A deep learning based algorithm for multi-criteria recommender systems. *Knowledge-Based Systems*, *211*, 106545.

[40] Rama, K., Kumar, P. & Bhasker, B. Deep autoencoders for feature learning with embeddings for recommendations: a novel recommender system solution. *Neural Comput & Applic* 33, 14167–14177 (2021). https://doi.org/10.1007/s00521-021-06065-9

[41] M. Ahmadian, M. Ahmadi, S. Ahmadian, S. M. Jafar Jalali, A. Khosravi and S. Nahavandi, "Integration of Deep Sparse Autoencoder and Particle Swarm Optimization to Develop a Recommender System," *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2021, pp. 2524-2530, doi: 10.1109/SMC52423.2021.9658926.

[42] Yu, Miao & Tianqi, Quan & Peng, Qinglong & Yu, Xu & Liu, Lei. (2022). A model-based collaborate filtering algorithm based on stacked AutoEncoder. Neural Computing and Applications. 34. 10.1007/s00521-021-05933-8.

[43] James Kennedy and Russell C. Eberhart. *Swarm Intelligence.* Morgan Kaufmann Publishers, San Francisco, California, 2001.

[44] Ms. Richa Agnihotri, Dr. Shikha Agrawal, Dr. Rajeev Pandey, 2016, Multi Objective Particle Swarm Optimization: A Survey, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 05, Issue 02 (February 2016), http://dx.doi.org/10.17577/IJERTV5IS020467

[45] Rahul Katarya & Om Prakash Verma. (2018). Efficient music recommender system using context graph and particle swarm. Multimedia Tools and Applications, Volume 77, Issue 2, pp 2673–2687.

[46] Havens TC, Bezdek JC, Leckie C et al (2012) Fuzzy c-means algorithms for very large data. IEEE Trans Fuzzy Syst 20:1130–1146

[47] Bezdek JC, Ehrlich R, Full W (1984) FCM: the fuzzy c-means clustering algorithm. Comput Geosci 10:191–203

e-Prime - Advances in Electrical Engineering, Electronics and Energy 2 (2022) 100076

Contents lists available at ScienceDirect

e-Prime - Advances in Electrical Engineering, Electronics and Energy

journal homepage: www.elsevier.com/locate/prime

A review on modelling and techniques used for piezoelectric power generation from vibration of geo-structures

Nisha Kumari *, Ashutosh Trivedi

Civil Engineering Department, Delhi Technological University, Delhi 110042, India

ARTICLE INFO

Keywords:
Structures
Vibration, Geo-structure
Piezoelectric power generation
Power output

ABSTRACT

The engineering structures consist of machine foundations, pavement, subgrade, bridge abutment, railway track, retaining wall, and contact geo-material. These structures are subjected to ambient mechanical vibration due to dynamic loads. The mechanical vibration of the structures can be utilized for power generation using the electromechanical conversion characteristics of the piezoelectric material. This paper captures recent advances in modelling *PEH* (piezoelectric energy harvester) and numerical techniques for power generation from structures. The electromechanical coupling of the structures with *PEH* is the key factor for the voltage and power output. The coupling techniques of structures and energy harvesters are significantly affected by the variables, namely fundamental frequency, the contact pressure of loading, the geometry of the structure, moving load frequency, placement of harvester, and dynamic response of the contact material. The mode of operation of *PEH* (31 vs. 33 modes) significantly influences the dynamic input and power output from the structure-power generation system. The present study demonstrates the potential of power generation from mechanical vibration of geo-structures has been classified as low, moderate, and high power output. It recommends the coupling parameters for the structural design to improve the efficacy of power generation from the ambient vibration of geo-structures.

Introduction

The engineering structures contribute to the social and economic development of the nation. Examples of engineering structures are machine foundations, pavements subgrade, bridge abutments, railway tracks, and retaining walls in contact with the supporting geomaterial. These geo-structures are subjected to ambient mechanical vibrations due to various dynamic loads. These vibrations are being wasted as mechanical energy. The researchers are fascinated with collecting this mechanical energy and converting it into electrical energy [1,2]. This unused energy can harvest using specific techniques and provide sustainable power sources in remote areas, transport safety, structural health monitoring, and traffic management [3–8]. According to a report, worldwide power utilization is predicted to increase by 44% from 2006 to 2030 [9,10]. Based on the literature, three conversion techniques are used: electromagnetic, piezoelectric, and electrostatic. While each technique can generate a significant amount of power, *PEHs* systems have received more attention because they receive the mechanical vibration directly from the system and convert it into electrical energy [11–13]. According to Chen et al. [14], piezoelectric polymers are flexible energy harvesters due to their soft nature. The flexibility of contact geomaterial utilises the property of piezoelectric polymers in power generation. As the performance of the PZT depends on the mechanical vibrations and stress- strain response. The durability of the PZT patches depends on the stress generated and the intensity of the heavy moving loads. The investigators performed laboratory tests to check the damage resistance of the piezoelectric material. It was observed that beyond a year or so the PZT patches may be replaced if not protected and serviced well before time [15].

The mechanical energy of vibrations of roads, bridges, and retaining structures can be converted by utilizing electromechanical conversion [16,17]. The conversion of mechanical energy depends upon the energy stored in the material using an implement consisting of PZT. Few researchers analyzed the feasibility of power generation from vehicle-induced vibration on bridges and pavement [3,18]. The power output of mechanical energy has been significantly influenced by the parameters of material properties, frequency of the vibration, and stress-strain characteristics of the material. A piezoelectric generator could generate 100mW power from railway track vibrations [19]. In the last decade, a considerable number of attempt has been made to generate significant power from the civil structures. Structural elements

* Corresponding author.
E-mail address: nishasoni.ce@gmail.com (N. Kumari).

https://doi.org/10.1016/j.prime.2022.100076
Received 23 June 2022; Received in revised form 8 September 2022; Accepted 7 October 2022
Available online 8 October 2022
2772-6711/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Notations

$\widetilde{K}$	added stiffness induced by the applied preloading conditions
$\dot{q}(t)$	the current output of piezoelectric materials
Θ	electromechanical coupling matrix
D	hysteretic damping ratio
C	coefficient of viscous damping
EI	flexural stiffness of the beam
$w(x,\ t)$	transverse deflection of the pavement
$\overline{w}(\overline{x},t)$	transverse deflection of the bridge
m_a,m_b	masses of asphalt mixtures
δ	Dirac delta function
t_d	time required to load pass the bridge
k_2	elastic coefficient of the stack
C_2	equivalent capacitance of the stack
N_s	electromechanical conversion coefficient of the stack
$y(x,t)$	displacement of the bridge at position x and time t
n	the ratio of modulus PZT over modulus backfill
μ	the ratio of thickness PZT over thickness backfill
a	the ratio of thickness retaining wall over thickness PZT
b	the ratio of modulus retaining wall over modulus PZT
u_x	horizontal displacement of retaining wall-PZT system
u_z	vertical displacement of retaining wall-PZT system
$w_{rel}(x,t)$	transverse displacement of the beam
$w_b(x,t)$	transverse displacement of the base
C_a	viscous air-damping
$V(t)$	steady-state voltage
R	external resistance
C_c	the capacitance of the piezoelectric material
ϑ	coefficient of background coupling terms
K_s	stiffness of compression spring
nd_{33}	the effective piezoelectric constant of the stack
nk_{33}^{σ}	the effective dielectric constant of the stack
E_z	electric field
ε_i	the linear strain of deformation
k_i^o	the curvature of the plate at the midplane
$\widehat{z}$	thickness matrix
B_u	strain-displacement matrix
N_u	displacement shape function
u^e	discrete element nodal displacement vector
P_3	piezoelectric polarization in 3-D
T_i	stress tensor
g_{3i}	piezoelectric voltage constant of PZT
ε_{33}^T	relative dielectric constant in 3-D
U_E	harvested electric energy from a PZT disk
I	the second moment of the area around the neutral axis
C_a	viscous damping coefficient
D	dynamic stiffness matrix
C_p	cost of PZT per unit
C_i	cost installation of PZT per unit
w_p	energy generated from each PZT per vehicle
N	number of vehicles passing per day
Y	the service life of the PZT
FEA	finite element analysis

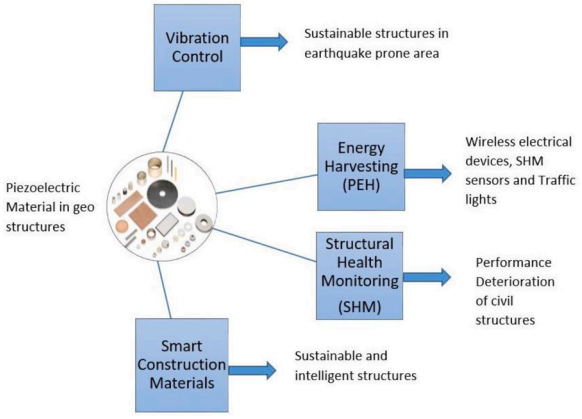


Fig. 1. Piezoelectric applications in structures utilizing mechanical vibrations.

and numerical methods are developed to convert wasted mechanical energy into electrical energy. The parameters, namely geometric design, orientation, moving loads, frequency, and dynamic response of the system considered [3,17,20,21]. Piezoelectric materials have been used for vibration control [22], structural health monitoring [23,24], and smart construction materials [25]. The vibration power generation technique is also utilized to reduce the noise in a delicate timber building [26]. The application of piezoelectric power harvesting in civil engineering is broadly explained in Fig 1. The power output can be utilized for wireless devices, structural health monitoring sensors [4,6, 27], hybrid electrical vehicles, and traffic lights [28].

The principal objective of this paper is to provide a comprehensive survey on the recent advances in modelling and numerical techniques for utilising mechanical vibration of geo-structures for piezoelectric power generation. The detailed research methodology of the review is shown in Fig. 2. The modelling of various *PEHs* used in the basements, retaining walls, pavements, bridges and railway tracks are discussed. The working principles and numerical techniques, namely, the finite element method and dynamic stiffness method are demonstrated for power generation from mechanical vibration of structures. This study focuses on several important aspects of using *PEHs* in civil engineering, including structural geo-materials, geometric design and numerical methods. Different *PEHs* are compared in terms of power output from the granular subcases, pavements, bridges and railway tracks. The advantages and limitations of piezoelectric power generation modelling and techniques are presented. The future research recommendations of piezoelectric power generation from ambient vibration of geo-structures are discussed.

Structural elements and modelling of PEH

This section describes the advancement of the structural elements and modelling the energy harvesters for utilising the mechanical vibrations from structures. Many researchers proposed the different structural elements of the harvester for conversion of mechanical vibration of structures namely, bridges [28–31], and pavements [17, 32–35], retaining walls [36,37]. The structural element modelling was designed based on the stress-strain response of the structures.

The selection of the structural element for the energy harvester has been made based on the frequency of the mechanical vibrations. To harvest the energy from the low-frequency vibrations stress-based harvesters are used as shown in Fig 3. The stacks and cymbal structure elements are most widely used in converting low-frequency vibration from structures [38]. The piezoelectric cymbal transducer provides a potential power generation under environmental dynamic loads such as moving loads on roadways pavements. Due to stiffness consistency and energy conversion efficiency, the cymbal and bridge transducer design has been recommended. The Cymbal piezoelectric transducers can generate power (1.2mW) at a 20Hz frequency for each passing vehicle

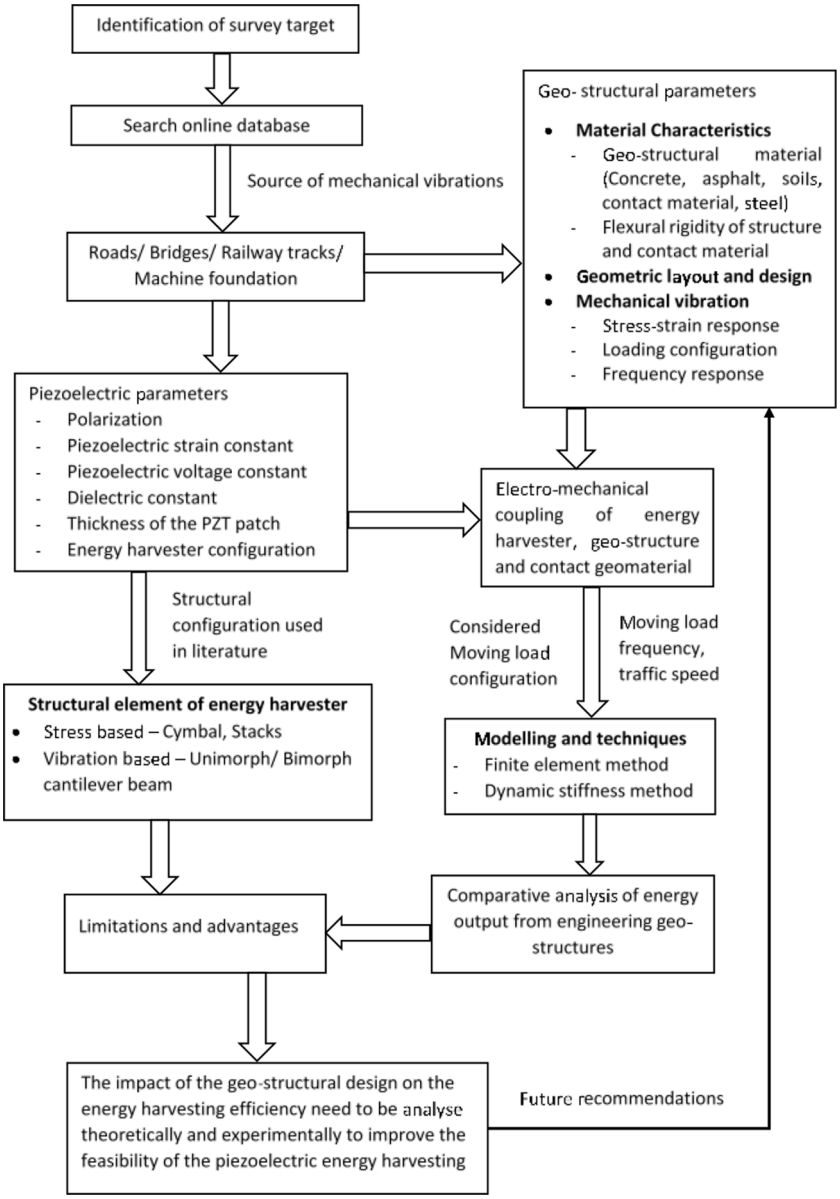


Fig. 2. Flow chart for power generation from mechanical vibration of the geo-structure.

and the electrical voltage generated around 97.33 V (0.06J) per vehicle from asphalt pavement. The deformation of the pavement due to mechanical vibration changes the coupling of cymbal transducers and affects the voltage generation [35,39]. The maximum 65MWh power could be produced from a 100 m roadway with around 30,000 piezoelectric transducers in a year [18].

In a piezoelectric cymbal transducer-based energy harvester, the direct piezoelectric effect is considered, and polarization occurs on the vertical surface of the PZT material. The mathematical relation between piezoelectric polarization of the cymbal transducer and stress due to mechanical vibration [39] is expressed as,

$$P_3 = \sum_{i=1}^{i=6} d_{3i} T_i \tag{1}$$

where P_3 is the piezoelectric polarisation in the 3rd principal direction, $d_{3i}$is the piezoelectric strain constant, and T_i is the stress tensor.

The initial electric field generated in PZT can express as,

$$E_3 = \frac{P_3}{\varepsilon_{33}^T} \text{ or } E_3 = \sum_{i=1}^{i=6} g_{3i} T_i \tag{2}$$

The PZT patch polarization generates the voltage that can be calculated from the following equation,

$$V_3 = \int E_3 dt_p = \sum_{i=1}^{i=6} g_{31} T_i dt_p \tag{3}$$

where $g_{3i}$is the piezoelectric voltage constant of the PZT$\left(g_{3i} = \frac{d_{3i}}{\varepsilon_{33}^T}\right)$, $t_p$is the thickness of the PZT and ε_{33}^T is the relative dielectric constant in the third principal direction. The power output of the piezoelectric cymbal transducer-based energy harvester is expressed as,

$$U_E = \frac{1}{2} P_3 E_3 A t_p = \frac{1}{2} V^2 \frac{\varepsilon_{33}^T \varepsilon_o A}{t_p} \tag{4}$$

where, U_E is the harvested electric energy, and A is the surface area of the PZT patch.

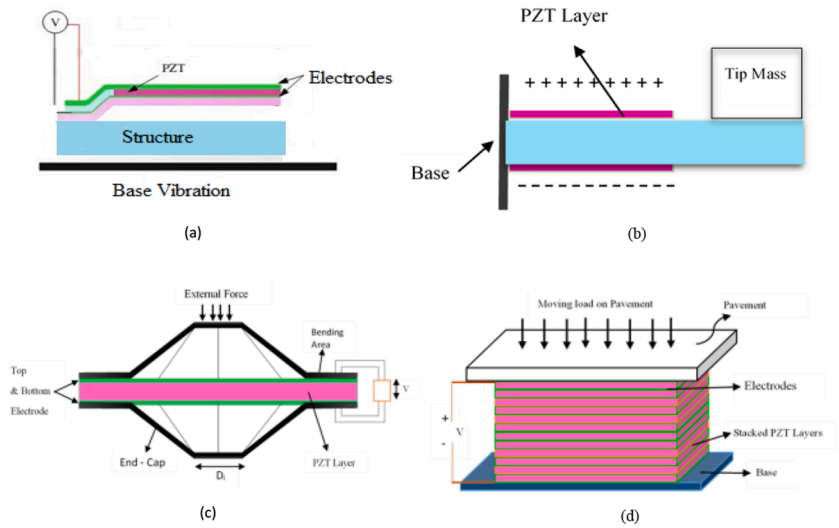


Fig. 3. The piezoelectric elemental structures (a) Unimorph piezoelectric cantilever beam; (b) Bimorph piezoelectric cantilever beam with tip mass; (c) The design of cymbal transducer made of PZT material; (d) The geometry of piezoelectric stack for structure.

The mechanical and electrical behaviour of piezoelectric stack embedded in asphalt pavement found excellent power sources using the finite element method [17]. The stack piezoelectric devices installed on the railway track could harvest energy up to 1000mJ when the stiffness increased significantly (MN/m) [40]. The piezoelectric stack transducer consists of a displacement sensor, a compression spring, a piezoelectric stack, and force transmission units. The inertial force in the transducer is neglected for simplification. Mathematical modelling has been proposed to analyze the voltage output [41]. The vertical stress due to moving loads is expressed as,

$$\sigma_Z = q(t) = \frac{K_s \omega}{A} \tag{5}$$

Where K_s is the stiffness of the compression spring, ω is the excitation frequency, and A is the contact area of loading. The stack-type harvesters consist of multiple no. of piezoelectric patches. Thus the constitutive equation of the piezoelectric layers is expressed as,

$$D_Z = nd_{33}\sigma_z + n\varepsilon_{33}^{\sigma} E_z \tag{6}$$

where nd_{33} and nk_{33}^{σ} are the effective piezoelectric strain and dielectric constants of the stack, which depend upon stresses due to the mechanical vibrations, respectively. D_z is the electrical displacement in the vertical direction. The voltage across the resistance is expressed as,

$$\frac{d}{dt}\left(\int_A D_z dA\right) = \frac{V(t)}{R} \tag{7}$$

where A is the area of the piezoelectric element. For the harmonic vibrations [$q(t) = q_o e^{j\omega t}$], the complex voltage is expressed as,

$$V_o = \frac{j\omega n d_{33} A q_o}{j\omega C_p + (1/R)} \tag{8}$$

The capacitance (C_P) is the component of the electrical circuit. The steady-state voltage is calculated using$V(t) = V_o e^{j\omega t}$.

The vibration from the bridge and railway tracks is converted into electrical energy using vibration-based energy harvesters. The vibration-based energy harvesters consist of a layer of piezoelectric material with a flexible cantilever beam with tip mass as shown in Fig. 3 (b). The change in the thickness of the PZT patch, tip mass and loading conditions influence the power output from the mechanical vibrations [42, 43]. The 0.53mW of power had been generated from one harvester at a load resistance of 60KΩ from bridge vibrations[44]. The unimorph energy harvester could generate about 13.8V under cyclic loading of 10kN [45]. The moving vehicle modelling and surface conditions change the coupling effect of the piezoelectric cantilever energy harvester significantly [46]. A Bimorph Cantilever beam with a thin PZT patch could produce 1mW of power at 25Hz [47].

There are several models proposed for a piezoelectric cantilever-based harvester. For the piezoelectric dynamic system [48,49]. The electromechanical coupling effect of the cantilever energy harvester is based on the Euler-Bernoulli beam theory and Gaussian law. The governing equation of motion and voltage output for the cantilever-based bimorph energy harvester for the transverse forced vibrations for the external damping is expressed as,

$$EI\frac{\partial^4 w_{rel}(x,t)}{\partial x^4} + C_S I\frac{\partial^5 w_{rel}(x,t)}{\partial x^4 \partial t} + C_a\frac{\partial w_{rel}(x,t)}{\partial t} + m\frac{\partial^2 w_{rel}(x,t)}{\partial t^2} + \vartheta V(t)\left[\frac{d\delta(x)}{dx} - \frac{d\delta(x-L)}{dx}\right] = [-m + M_t\delta(X-L)]\frac{\partial^2 w_b(x,t)}{\partial t^2} - C_a\frac{\partial w_b(x,t)}{\partial t} \tag{9}$$

$$\frac{V(t)}{R} + C\frac{dV(t)}{dt} + \vartheta\int_0^L \frac{\partial^3 w_{rel}(x,t)}{\partial x^2 \partial t}dx = 0 \tag{10}$$

where, the bending stiffness is EI, $w_{rel}(x,t)$and $w_b(x,t)$are the transverse displacement of beam and base, respectively, C_a is the viscous air-damping coefficient, m is the mass per unit length and C_S is the strain-rate damping coefficient. $V(t)$ is the voltage in the circle, C is the internal capacitance of the piezoelectric layer, R is the external resistance,

M_t is the participating mass, and ϑ is the coefficient of the background coupling terms.

Numerical techniques

Past research showed that ambient vibrations are a good power source (50 -250μW/cm^3) due to their easy accessibility in structural fields [17,50,51]. Various methods have been used to understand the feasibility of piezoelectric power generation from the mechanical vibration of the structure and its ability to produce significant power. The two types of methods finite element method and dynamic stiffness method are presented to analyse the dynamic response of the piezoelectric structure and geo-structures.

Piezoelectric finite element method

The structure analysis of embedded or bonded piezoelectric sensors requires the precise electromechanical modelling of mechanical and electrical responses. The finite element model for piezoelectric power generation has been utilized to discretize the electromechanical phenomenon. The finite element method of piezoelectric material is widely used for power generation from dynamic loads [18]. The structure is divided into finite numbers of non-overlapping elements in the finite element analysis to calculate the unknowns at different nodes. The mechanical displacement (u), applied force (f) and electrical potential are evaluated using nodal values of mechanical displacement $\{u(x_i, y_i)\}$and the electric potential$\{\phi(x_i,y_i)\}$. The corresponding displacement nodal shape function can be defined as a potential nodal shape function$[N_\phi(x_i,y_i)]$. For the n no. of nodes, the mechanical displacement function has been evaluated using the following equations:

$$\{u(x,y\} = [N_u(x_i,y_i)]\{u(x_i,y_i\} \tag{11}$$

$$\{\phi(x,y\} = \left[N_\phi(x_i,y_i)\right]\{\phi(x_i,y_i\} \tag{12}$$

Therefore, using the constitutive equation of piezoelectric material the strain $\{\varepsilon\}$ and electric field $\{E\}$ is defined as,

$$\{\varepsilon\} = [B_u]\{u(x_i,y_i\} \tag{13}$$

$$\{E\} = -\left[B_\phi\right]\{\phi(x_i,y_i\} \tag{14}$$

Substituting the strain–displacement and electric potential in the equation of motion, the couple electromechanical has been obtained for the piezoelectric finite model as,

$$\{M\}\left\{\ddot{u}_i\right\} + [K_{uu}]\{u_i\} + \left[K_{u\phi}\right]\{\phi_i\} = \{F(x_i,y_i)\} \tag{15}$$

$$\left[K_{\phi u}\right]\{u_i\} + \left[K_{\phi\phi}\right]\{\phi_i\} = \{Q_i\} \tag{16}$$

where, $\{M\}$is the dynamically consistent mass matrix, $[K_{uu}]$ is stiffness matrix, $[K_{u\phi}]$ is piezoelectric stiffness matrix, and $[K_{\phi\phi}]$ is dielectric stiffness matrix.

The finite element model provides the power output efficiency of the *PEH* [52,53]. The finite element method is used to analyse the deformation, dynamic load response, and structural parameters of the structures namely, pavement, bridge and railway tracks [1,54–57]. The displacement profile of the system is significantly affected by the stiffness matrix. For the advancement in the modelling of dynamically loaded structures, the material stiffness of the different layers of the system is requisite. For geo-structures, the stiffness matrix can be evaluated from the equivalent stiffness of the different layers of the system. By implementing this concept Eq. (15) can be modified as follows,

$$\{M\}\left\{\ddot{u}_i\right\} + \left[K_{eq(uu)}\right]\{u_i\} + \left[K_{u\phi}\right]\{\phi_i\} = \{F(x_i,y_i)\} \tag{17}$$

$[K_{eq(uu)}]$is the equivalent stiffness matrix for the displacement in multiple layers of the geo-structure.

Table. 1
Comparison of output potential difference as per harvesting technologies for substructures.

Type of Piezoelectric harvester	Stress (MPa)	Potential difference (V)	Power output (mW)
Bridge transducer [a]	0.7	556	28.7
Static loading prototype containing four PZT disks [b]	5	650	-
Composite piezoelectric transducer [c]	0.21, 0.35, 0.71	22, 36, 70	-
Ball-type protective pad with transducers [d]	0.7, 0.2	-	50.41, 2.92
Cymbal piezoelectric transducer [e]	0.7	-	1.2

[a] Jasim et al. [61]
[b] Roshani et al. [60, 64]
[c] Xu et al. [33]
[d] Wang et al. [65]
[e] Zhao et al. [39]

Dynamic stiffness method

The dynamic stiffness method is based on the exact solution of the wave equation. The finite element method is more complex than the dynamic stiffness method. It works on the concept of frequency response functions. The dynamic stiffness method is utilized to model beams with arbitrary boundary conditions. A dynamic stiffness method was proposed for modelling power generation beams based on the exact solution of the wave equation of motion [58]. The equation of motion bimorph piezoelectric harvester is expressed as,

$$B\frac{\partial^4 u}{\partial x^4} + A\frac{\partial^5 u}{\partial x^4 \partial t} + C_a\frac{\partial u}{\partial t} + m\frac{\partial^2 u}{\partial t^2} = 0 \tag{18}$$

$$B = Y_pI_p + Y_{sh}I_{sh}; A = c_pI_p + c_{sh}I_{sh} \tag{19}$$

where C_a is the viscous damping coefficient, u is the displacement, m is the mass per unit length, B is the bending stiffness of the bimorph energy harvester, and A is the structural damping. The dynamic stiffness method assumed the harmonic excitation. The harmonic excitation of the system can be expressed as,

$$u(x,t) = \mathrm{Re}\left\{\overline{u}(x)e^{iwt}\right\} \tag{20}$$

The dynamic stiffness matrix D of the beam with tip mass is expressed as,

$$f = Du \tag{21}$$

where,$f = [\overline{F}_o \quad \overline{\Gamma}_0 \quad \overline{F}_l \quad \overline{\Gamma}_l]^T$ and $u = [\overline{u}_o \quad \overline{\theta}_0 \quad \overline{u}_l \quad \overline{\theta}_l]^T$

The dynamic stiffness D_e of the beam without tip mass is expressed as,

$$f_e = D_eu \tag{22}$$

where $f_e = [\overline{F}_o \quad \overline{\Gamma}_0 \quad \overline{F}_{le} \quad \overline{\Gamma}_{le}]^T$and $\overline{u}$ and $\overline{\theta}$are the complex amplitude of displacement and rotation of the cantilever beam energy harvester, f and u are vectors of the complex amplitude and displacement, respectively. The dynamic stiffness method allows structures embedded with piezoelectric patches to respond in arbitrary high-frequency frequencies [59].

Power generation from mechanical vibration

Mechanical vibration of pavement and subgrade

This section provides the application of different energy harvesters to generate the electrical energy from the mechanical vibration of the

Table. 2
The analytical techniques for the conversion of mechanical vibrations to electrical energy from structures namely pavements, subgrade, and retaining structures.

Technique	Input	Output
Vibration control of flexible retention systems [a]	The generalized equation of the spring-mass system $m\ddot{u} + k_{hp}u + k_{hs}u = F\sin(\omega t + \beta)$	Horizontal and vertical displacement amplitude and effect of the stiffness of the structure
The deformation of pavement system subjected to moving [b]	The response of the pavement system $w(x,t) = \frac{P}{2\pi}\int_{-\infty}^{\infty}\frac{\sin(r_0\xi)}{r_0\xi}\times \frac{e^{i\xi(x-vt)}}{EI\xi^4 + k(1+2iD) - \rho(\nu\xi)^2}$	Parametric study to understand the effect of damping and velocity of vehicles on the power output
Cymbal piezoelectric harvester embedded in asphalt pavement [c]	The electric energy stored in the Cymbal $U_E = \frac{1}{2}P_3E_3At_P = \frac{1}{2}V^2\frac{\varepsilon_r^T\varepsilon_0 A}{t_P}$ The displacement difference percent $pd = \frac{d_{wo} - d_w}{dw_o}\times 100$	The efficiency and coupling effects with a pavement of Cymbal with various sizes
The transformation of asphalt layers into a *PEH* using the 3-D electromechanical model and Cymbal PZT [d]	The equation of three-degree of freedom electromechanical model $m_a\ddot{x}_a + k_a(x_a - x_p - x_b) + c_a(\dot{x}_a - \dot{x}_p - \dot{x}_b) = Q_0\mathrm{Cos}(wt)$ $m_p\ddot{x}_p + nk_p(x_p - x_b) + nPV_p = Q_0\mathrm{Cos}(wt) - m_a\ddot{x}_a$ $nk_p(x_p - x_b) + nPV_p = m_b\ddot{x}_b + c_b\dot{x}_b + k_bx_b$ $-nP\dot{x}_p + nC_pV_p + \frac{V_p}{R} = 0$	The voltage outputs are analysed for different material properties, stiffness, shear modulus, piezoelectric layer design, and external dynamic loading conditions
Roadway energy harvester contains piezoelectric multilayer stacks [e]	The constitutive equations of the piezoelectric stack $\begin{bmatrix}F_p\\Q\end{bmatrix} = \begin{bmatrix}k_2 & -N\\N & C_2\end{bmatrix}\begin{bmatrix}x_2\\V_p\end{bmatrix}$	The harvester can generate sufficient energy for driving the equipment used in transportation infrastructure. K_2 represents the elastic coefficient

[a] Kumari and Trivedi [22]
[b] Xiang et al. [15]
[c] Zhao et al. [39]
[d] Guo and Lu [52]
[e] Jiang et al. [66]

pavement due to dynamic load. The coupling of the *PEH* and the pavements significantly affects the power output. The polarization of the piezoelectric material is directly proportional to the mechanical stress of the pavement and subgrade due to moving loads [Eq. 1]. Table 1 summarises the comparative voltage generation from the applied force of moving vehicles on the structures. The electrical potential is sensitive to the vertical load on the pavement, frequency and no. of the piezoelectric disk [60]. The moving loads due to heavy vehicles are considered to increase the power generation efficiency from the pavement and subgrade due to higher stress [36,37]. The Cymbal piezoelectric transducers can generate power (1.2mW) at a 20Hz frequency for each passing vehicle and the electrical voltage generated around 97.33 V (0.06J) per vehicle [35,39]. The cymbal piezoelectric harvester generates 16μW power for the passage of each heavy vehicle from asphalt pavements [18]. The piezoelectric stacks are suitable for high loading conditions, low frequency, and compatibility with pavements [17].

Moreover, the bridge transducer for power generation and produced a significant electrical potential (556V) for 0.7 MPa applied stress [61]. The increased number of piezoelectric elements in the power generation pavement system enhances the power output. Approximately 65MWh power could be produced from a 100 m roadway with around 30,000 piezoelectric transducers in a year [18]. The output power and power conversion efficiency of a piezoelectric transducer are determined by the geometric design and the placing in the pavement and subgrade. The embedment depth affects the magnitude of stress on the energy harvester, which vitally affects the power output from airfield pavements [62,63]. The effect of the stiffness of geo-structure and PEH on the electromechanical coupling can be analysed using the varied analytical techniques shown in Table 2.

Mechanical vibration of railway tracks, bridges, and retaining structures

Railway tracks, bridges, and railway tracks are the ambient vibration sources in civil structures. To harvest the piezoelectric energy, the analysis of the displacement due to moving loads of the railway tracks, bridge, and retaining structures [22,67]. Various piezoelectric structural elements have been utilized to enhance the power output from the mechanical vibration of the bridge. The vibration-based cantilever energy harvesters are most suitable due to the electromechanical coupling

Table. 3
Analytical techniques used for power generation applicable to geo-structures.

Technique	Input	Output
Piezoelectric cantilever beam harvester [a]	$EI\frac{\partial^4 w_{rel}(x,t)}{\partial x^4} + C_SI\frac{\partial^5 w_{rel}(x,t)}{\partial x^4\partial t} + C_a\frac{\partial w_{rel}(x,t)}{\partial t} + m\frac{\partial^2 w_{rel}(x,t)}{\partial t^2} + \vartheta V(t)\left[\frac{d\delta(x)}{dx} - \frac{d\delta(x-L)}{dx}\right] = [-m + M_t\delta(X-L)]\frac{\partial^2 w_b(x,t)}{\partial t^2} - C_a\frac{\partial w_b(x,t)}{\partial t}$	Parametric study of cantilever *PEH*
A bimorph *PEH* subjected to permanent axial loading [b]	The generalized equation of the system $M_T\ddot{r}(t) + C\dot{r}(t) + (K_T + \widetilde{K})r(t) - \Theta v(t) = 0$ $\Theta^T\dot{r}(t) + C_P\dot{v}(t) = \dot{q}(t)$	The effect of the pre-stress parameters on the power generation
Bridge vibrations due to moving load, moving mass and moving oscillator [c]	Moving load model, the load on the bridge $p(\overline{x},t) = \begin{cases}P\delta(\overline{x} - vt), & 0 \le t \le t_d\\ 0, & t \ge t_d\end{cases}$ Moving mass model, the load on the bridge $p(\overline{x},t) = \begin{cases}M\{g - [\ddot{\overline{w}}(\overline{x},t) + \ddot{\eta}(vt)]\}\delta(\overline{x} - vt), & 0 \le t \le t_d\\ 0, & t \ge t_d\end{cases}$	More power can be generated with consideration of the unevenness of the surface. The unevenness effect does not apply to the moving load model
Power generation from a highway bridge vibration from a moving point load [d]	The dynamic response of the bridge with a point load, P and velocity, $um\frac{\partial^2 y(x,t)}{\partial t^2} + c\frac{\partial y(x,t)}{\delta t} + EI\frac{\partial^4 y(x,t)}{\partial x^4} = \delta(x - ut)P$	Parametric study to obtain the power output and design the energy management system

[a] Erturk and Inman [74]
[b] Rhimi and Lajnef [75]
[c] Zhang et al. [42]
[d] Ali et al. [29]

Table. 4
The power output from mechanical vibrations of traffic loads.

Type of Model	Piezoelectric properties Elastic stiffness of PZT, (GPa) *Ep*	Piezoelectric strain constant (C/m^2) d_{31}	Piezoelectric strain constant (C/m^2) d_{33}	Density of PZT (Kg/m^3)	Power (mW)
Cantilever *PEH* (Bridge) [a]	56	10.4	-	7450	400
Piezoelectric stack transducer (Railway Track) [b]	62.22	-186	700	7450	0.15
Finite element PZT model (Train-Bridge) [c]	71	-274	-	7450	0.4

[a] Zhang et al. [42]
[b] Wang et al. [40]
[c] Cahill et al. [77]

effect. The cantilever beam creates a higher strain in the harvester for low-intensity ambient vibrations. A bimorph cantilever beam with a thin PZT patch could produce 1mW of power at 25Hz [47]. The unimorph energy harvester could generate about 13.8V under cyclic loading of 10kN from a steel-beam slab bridge [45]. To generate a high electric effect, load amplitude, loading frequency and the resultant strain should be considered and its piezoelectric effect can be assessed based on the generated electric effect per unit strain based on the coupling coefficient.

The enhanced frequency of non-linear energy harvesters improves power generation efficiency [68–70]. The dynamic non-linear response of low-frequency vibrations is based on mechanical and electrical energy coupling [71]. To obtain the maximum power from bridge vibration the resonant frequency must match the resonant frequency of the harvester [47,72]. The multi-impact energy harvesters with a natural frequency of fewer than 10 Hz are used for the low-frequency vibrations of the bridge [24,73]. The high-frequency piezoelectric cantilever harvester showed better results than low-frequency vibrations.

Table 3 shows the power generated from the various piezoelectric harvester: cantilever energy harvester, Cymbal piezoelectric transducer, stack transducer, and finite element model. It is observed that using the cantilever energy harvester highest power can be generated compared to others.

The ambient vibrations from railway tracks and tunnels are used for sensors embedded to monitor structural health [76]. The railway track vibrations at maximum load can be converted into voltage 50-70V and power of 100mW for wireless sensors [19]. Passenger trains produced more power output (0.59mW) than freight trains 0.03mW [77]. The geometric layout of the structure and installation position significantly affects the power output efficiency of the energy harvester [78]. Each piezoelectric patch in a stack located at the bottom of a steel rail could harvest energy up to 1000mJ when the stiffness increased significantly [40]. It is observed from the literature that dynamic response parameters, namely acceleration, frequency, and damping of moving loads, significantly affect the piezoelectricity and power output, as shown in Table 4.

Composition of PZT and construction material

This section analyses the potential application of piezoelectric material in health monitoring and vibration control. The property of PZT material allows possibilities for implementing sustainability through vibration control, structural health monitoring and smart construction materials [25,79]. The cement-PZT composites moderately produce the

Table. 5
Composites of PZT and construction material for power generation.

Material composition	Piezoelectric property d_{33} (pC/N)	Polarization Voltage (kV/mm)
Cement-PZT- composite [a,b,c,d]	29-71 20-80 110-215	1.5 1 6
PZT-PVDF-cement composite [e]	15-25	1
Silica fume –cement –PZT composite [f]	18-20	2
Carbon black-cement –PZT composite [g]	2-30	4
Aluminium-cement –PZT composite [h,i]	0.5-7	0.6-1
Asphalt – PZT composite [j]	287	4-7.6
Geo-structures (granular fill and PZT composite) [k]	310-500	-

[a] Gong and Zhang [94]
[b] Wang et al. [95] [c] Chaipanich and Jaitanong [96]
[d] Lam and Chan [97]
[e] Jaitanong et al. [98]
[f] Chainpanich [99]
[g] Gong et al. [84]
[h] Banerjee and Chennault [100]
[i] [101]
[j] Guo and Lu [52]
[k] Kumari and Trivedi [36]

Table 6
The advantages and limitations of the piezoelectric power generation technologies from geo-structures.

Technology	Advantages	Limitations
Piezoelectric cantilever beam harvester for bridge and railway tracks	An easy and simple method to calculate the strain under the base excitation in vibration.	Low energy conversion efficiency
Stack type piezoelectric transducer model for pavement and bridge	It is favourable for higher load applications. It is widely used for power generation from pavement and bridges.	The complex stacking process in limiting areas
Piezoelectric finite element model for bridge, railway track and pavement	Accurate prediction of mechanical displacement and electric potential	Less accurate for the static response for the isotropic piezoelectric material
Cymbal piezoelectric modelling equations for pavement	Applicable to harvest the energy from low-frequency loading.	Not applicable for high resonant frequency models
Dynamic stiffness method	Works with force and displacement frequency response function	The inclusion of damping is challenging
Dynamic response analysis for retaining wall	To understand the mechanical displacement at a low frequency	Not applicable for non-linear cross-sections

electrical current with the application of compressive force [80,81]. The electromechanical impedance technique is used for monitoring the cement compositions using piezoelectric materials, which is the function of the density and stiffness of the material [57,79,82,83]. The geo-structural characteristics significantly affect the stress-strain behaviour [36], which directly changes the impedance of the piezoelectric materials. The impedance of the PZT was analysed by changing the properties of the composite materials [5]. The electromechanical effect of the PZT material is influenced by the piezoelectric coefficient (d_{33}), which depends on factors namely polarization, temperature, and fabrication [84]. The researchers focused on the damage detection of concrete structures, steel structures and bridges using piezoelectric-based sensors [6,7,85]. The mechanical behaviour of the PZT-cement composite is significantly affected by the material properties namely porosity and young's modulus, participating mass, Poisson's

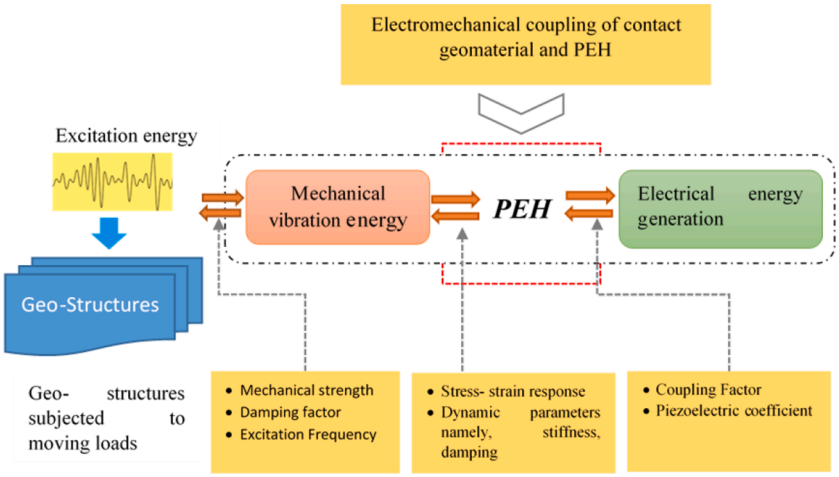


Fig. 4. The electromechanical coupling of contact geomaterial and *PEH* for power generation from mechanical vibration of geo-structures.

Table 7
Voltage and power output from varied geo- engineering infrastructures using PEH.

Geo-structure	Type of *PEH*	Properties	Voltage Output (V)	Power Output (mW)
Bridge [a]	Cantilever *PEH*	Even surface Uneven surface	-	0.0008 – 0.0015 30-500
Rail Bridge [b]	Cantilever *PEH*	Natural frequency Acceleration	0.03 – 0.4	-
Pavement [c]	Multilayer stack *PEH*	Traffic speed, type of vehicle	-	136000 -720000
Bridge[d]	Bimorph *PEH*	Vehicle speed		0.5-0.75
Railway bridge [e]	Unimorph *PEH*	Single span Double span	-	0.0091-462 0.008-290
Pavement [f]	Composition of asphalt and PZT	Vibration frequency (30Hz)	-	1.2- 300
Pavement [g]	Multilayer stack *PEH*	Vehicle speed and traffic volume resistance	-	1000-20000003-90
Pavement [h]	Cymbal *PEH*	Contact pressure		0.06-1.2
Bridge [i]	Unimorph *PEH*	Loading frequency and amplitude	4.9 -57	-
Pavement [j]	Stacked *PEH*	Loading frequency	20- 70	-
Bridge and Pavement [k]	Cantilever *PEH*	Acceleration	-	1-9

[a] Zhang et al. [42]
[b] Cahill et al. [77]
[c] Roshani et al. [102]
[d] Galchev et al. [24]
[e] Song [53]
[f] Guo et al. [52]
[g] Jiang et al. [66]
[h] Zhao et al. [39]
[i] Kim et al. [45]
[k] Ashebo et al.[3]

ratio and energy transfer function [63,86,87]. The polarization and electric admittance of the PZT material affect the potential of its application in structures [88–92].

The geo-structure sustainability can be improved by vibration control due to the dynamic loads. According to the concept of structure vibration control by adjusting the stiffness and mass parameter control

Table 8
Classification of power output from mechanical vibrations of geo-structure.

Source of mechanical vibration	Power Output		
	Low (mW)	Moderate (W)	High (W)
Structures (coupled with contact geo-material)	0.006 - 300	0.3-100	110-2000

of the dynamic variables such as displacement, velocity, and acceleration [93]. Consistent with it, a solution is developed to control the vibration of the retaining structure using the semi-active control strategy, considering the stress-strain response of the PZT patch [36]. Table 5 shows the composition of piezoelectric materials with construction materials like concrete, cement, and asphalt. The composite material concept promotes piezoelectricity and enhances the power generation capacity of the material.

Advantages and limitations of the harvesting techniques

Table 6 shows the advantages and limitations of piezoelectric power generation technologies in geo-structures. Stacks and cymbal modelling can be utilized for heavy loads on the subgrade, pavements, retaining, and bridges. The dynamic stiffness method provides the displacement and frequency response *PEH* (piezoelectric energy harvester). In the case of retaining walls, the dynamic response analysis gives the displacement profile. The stress-strain response of the host geo-structure influences the harvested power from *PEH*.

Voltage and power output from engineering geo-structures

The moving loads excite transients in the geo-structures, and the dynamic response is likely to correspond to the vibrational modes of structures. Thus, the material and geometric parameters of structures (pavement, bridge, retaining wall, and railway track) affect the magnitude of generated power. The precise power magnitude of power can be obtained through the dynamic response analysis of structural vibration. The factors influencing PEH from the mechanical vibrations of the geo-structures are shown in Fig. 4.

Different PEH systems were used to harvest energy from vibration due to moving loads: piezoelectric cantilever beam, the bimorph *PEH*, cymbal transducers, and piezoelectric multilayer stacks. Each vibration energy harvester has a predominant resonance frequency. Cymbal transducers and piezoelectric stacks energy harvesters improve power generation due to the high stiffness and natural frequency. The Cymbal and stakes transducers can withstand more considerable mechanical

stress than the cantilever energy harvesters can because the fundamental bending mode of a cantilever is much lower than the other vibration modes. Table 7 summarizes the power output from the road bridges, pavements, and railway bridges with different *PEHs*.

The power output from the structure based on the electromechanical coupling with the contact geo-material is classified as low (wireless sensors), moderate (structural health monitoring and electronic devices) and high (traffic lights, charging of hybrid vehicles) power output as shown in Table 8.

It has been observed that power output is significantly affected by the parameters namely, contact geomaterial friction of the surface subjected to moving loads, vehicle types, vehicle speed, traffic volume, amplitude, and frequency of induced vibrations. Multilayer stack *PEH* in pavements with heavy traffic volume and speed of more than 30m/s can generate a maximum power of 2kW [66]. The proper alignment position of *PEH* near the surface increases the contact pressure of moving vehicles on the structural elements. The impact pressure of moving vehicles increases the power output from the pavements [39]. The multilayer stacks and cymbal *PEH* are relevantly capable of improving the current and voltage for low-frequency stress conditions. To generate the maximum energy from the mechanical vibrations, the natural frequency of the *PEH* should be resonant with the natural frequency of the structures. The energy conversion efficiency can be captured from the stress-strain and vibration characteristics of structures as classified in Table 8.

Conclusions

This study summarises the modelling methods and numerical techniques for utilising the mechanical vibration of geo-structures in piezoelectric power generation. The stress-based and vibration-based modelling of *PEHs* is used to convert the mechanical vibration of geo-structures into electrical energy. The finite element and dynamic stiffness techniques are most suitable for accurately predicting the mechanical displacement and coupling effect due to ambient vibration. The inclusion of damping prominently present due to contact geomaterial introduces analytical complexity in the dynamic stiffness method.

The coupling of geo-structures and energy harvester is significantly affected by variables namely fundamental frequency, the contact mechanism of loading, the geometry and material characteristics of the geo-structure, moving load frequency, and dynamic response. The composition of the PZT with geo-material increases polarization and generates a higher voltage. Applying a stiffer stack configuration for low-intensity vibration is an inappropriate choice to create higher strain in the harvester. The cantilever beam should be embedded in 31 modes to generate more energy.

The present study demonstrates the potential of power generation from mechanical vibration of geo-structures classified as low, moderate and high power output in the range of 0.006-300mW, 0.300-100W and 110-200W, respectively. The recommended coupling parameters for the structural design have been shown to improve the efficacy of power generation from the ambient vibration of geo-structures.

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

The financial support received under the research project entitled “Energy Harvesting through Mechanical Vibrations Generated in Retaining Structure” (grant F. No. DTU/IRD/619/2105) from Delhi Technological University, Delhi, is well acknowledged.

Appendix A. Mathematical description for stress-strain and voltage output of *PZT* patch

When piezoelectric materials are subjected to mechanical stress, electrical energy generated in proportion to that stress is known as the direct piezoelectric effect. The direct effect occurs through the compression of piezoelectric material. The piezoelectric describes the combined effect of electric and elastic mechanical behaviour.

The linear electrical behaviour of the PZT material is expressed as,

$$D_i = \varepsilon_{ij} E_j \tag{A.1}$$

Hooke’s law for elastic materials is expressed as,

$$S_{ij} = s_{ijkl}\sigma_{kl} \tag{A.2}$$

These relations can be combined into the coupled equation, expressed as the following constitutive piezoelectric equations in the function of various piezoelectric coefficients. The piezoelectric energy is captured as charge density and voltage generation on the surface of PZT patches due to mechanical vibrations. The relation between in stress-strain, charge density and voltage output is expressed as the following relations,

The Strain-charge form

$$S = s_E.T + d^t.E \tag{A.3}$$

$$D = d.T + \varepsilon_T.E \tag{A.4}$$

Stress-Charge form

$$T = c_E.S - e^t.E \tag{A.5}$$

$$D = e.S + \varepsilon_s.E \tag{A.6}$$

Strain-voltage form

$$S = s_D.T + g^t.D \tag{A.7}$$

$$E = -g.T + \varepsilon_T^{-1}.D \tag{A.8}$$

Stress-voltage form

$$T = c_D.S - q^t.D \tag{A.9}$$

$$E = -q.S + \varepsilon_s^{-1}.D \tag{A.10}$$

where T is the stress component. S is strain components. E is the electric field. D electric charge density displacement. s is compliance coefficients, c is stiffness coefficient, ε is electric permittivity, d is piezoelectric coupling coefficient for strain change form, e is piezoelectric coupling coefficient for stress change form, g is the piezoelectric coefficient for strain-voltage, and q is piezoelectric coupling coefficients for stress voltage.

References

[1] C.H. Sun, G.Q. Shang, Y.K. Zhang, J.H. Du, Designing piezoelectric harvesting unit from road vibration, Adv. Mater. Res. Trans. Tech. Publ. 712 (2013) 1368–1371.

[2] Y. Chen, H. Zhang, Y. Zhang, C. Li, Q. Yang, H. Zheng, C. Lü, Mechanical energy harvesting from road pavements under vehicular load using embedded piezoelectric elements, J. Appl. Mech. 83 (2016).

[3] D.B. Ashebo, C.A. Tan, J. Wang, G. Li, Feasibility of energy harvesting for powering wireless sensors in transportation infrastructure applications. Nondestructive characterization for composite materials, aerospace engineering, civil infrastructure, and homeland security 2008, Inter. Soc. Opt. Photon. 6934 (2008) 69340Y.

[4] N.G. Elvin, N. Lajnef, A.A. Elvin, Feasibility of structural monitoring with vibration powered sensors, Smart Mater. Struct. 15 (2006) 977–986.

[5] S. Bhalla, C.K. Soh, Structural health monitoring by piezo-impedance transducers. I: modeling, J. Aerospace Eng. 17 (2004) 154–165.

[6] C.K. Soh, K.K. Tseng, S. Bhalla, A. Gupta, Performance of smart piezoceramic patches in health monitoring of a RC bridge, Smart Mater. Struct. 9 (2000) 533.

[7] D. Ai, H. Zhu, H. Luo, J. Yang, An effective electromechanical impedance technique for steel structural health monitoring, Constr. Build. Mater. 73 (2014) 97–104.

[8] J.P. Lynch, An overview of wireless structural health monitoring for civil structures, Philosophical Transactions of the Royal Society A: Mathematical, Physical Engineering Sciences 365 (2007) 345–372.

[9] S. Ahmad, M. Abdul Mujeebu, M.A. Farooqi, Energy harvesting from pavements and roadways: a comprehensive review of technologies, materials, and challenges, Int. J. Energy Res. 43 (2019) 1974–2015.

[10] C. Williams, R.B. Yates, Analysis of a micro-electric generator for microsystems, Sens. Actuators A: Phys. 52 (1996) 8–11.

[11] S.R. Anton, H.A. Sodano, A review of power harvesting using piezoelectric materials (2003–2006), Smart Mater. Struct. 16 (2007) R1–R21.

[12] L. Li, J. Xu, J. Liu, F. Gao, Recent progress on piezoelectric energy harvesting: structures and materials, Adv. Compos. Mater. 1 (2018) 478–505.

[13] V. Singh, D. Meena, H. Sharma, A. Trivedi, B. Singh, Investigating the role of chalcogen atom in the piezoelectric performance of PVDF/TMDCs based flexible nanogenerator, Energy Environ. Sci. 239 (2022), 122125.

[14] Y. Chen, X.J.S. Xue, Advances in the structural health monitoring of bridges using piezoelectric transducers, Sensors 18 (2018) 4312.

[15] H.J. Xiang, J.J. Wang, Z.F. Shi, Z.W. Zhang, Theoretical analysis of piezoelectric energy harvesting from traffic induced deformation of pavements, Smart Mater. Struct. 22 (2013).

[16] H. Li, C. Tian, Z.D. Deng, Applied physics reviews, Appl. Phys. Lett. 113904 (2014) 104.

[17] J.-f. Lv, K. Yang, L. Sun, W.-h. Chen, Y.-q. Tan, Finite element analysis of piezoelectric stack transducer embedded in asphalt pavement, in: 2015 Symposium on Piezoelectricity, Acoustic Waves, and Device Applications (SPAWDA). IEEE, 2015, pp. 152–156.

[18] A. Moure, M.I. Rodríguez, S.H. Rueda, A. Gonzalo, F. Rubio-Marcos, D. U. Cuadros, A. Pérez-Lepe, J. Fernández, Feasible integration in asphalt of piezoelectric cymbals for vibration energy harvesting, Energy Convers. Manage. 112 (2016) 246–253.

[19] Y. Tianchen, Y. Jian, S. Ruigang, L. Xiaowei, Vibration energy harvesting system for railroad safety based on running vehicles, Smart Mater. Struct. 23 (2014), 125046.

[20] M. Karimi, A. Karimi, R. Tikani, S. Ziaei-Rad, Experimental and theoretical investigations on piezoelectric-based energy harvesting from bridge vibrations under travelling vehicles, Int. J. Mech. Sci. 119 (2016) 1–11.

[21] E. Varadha, S. Rajakumar, Performance improvement of piezoelectric materials in energy harvesting in recent days-a review, J. Vibroeng. 20 (2018).

[22] N. Kumari, A. Trivedi, Vibration Control of Flexible Retention Systems. Advance in Computer Methods Geomechinics. Lecture Notes in Civil Engineering, Springer, 2020, p. 56.

[23] R. Shanker, S. Bhalla, A. Gupta, M. Praveen Kumar, Dual use of PZT patches as sensors in global dynamic and local electromechanical impedance techniques for structural health monitoring, J. Intell. Mater. Syst. Struct. 22 (2011) 1841–1856.

[24] T. Galchev, J. McCullagh, R. Peterson, K. Najafi, Harvesting traffic-induced vibrations for structural health monitoring of bridges, J. Micromech. Microeng. 21 (2011), 104005.

[25] Z. Li, S. Huang, L. Qin, X. Cheng, An investigation on 1–3 cement based piezoelectric composites, Smart Mater. Struct. 16 (2007) 999.

[26] C. Wei, X. Jing, A comprehensive review on vibration energy harvesting: Modelling and realization, Renew. Sust. Energ. Rev. 74 (2017) 1–18.

[27] H. Gu, G. Song, H. Dhonde, Y. Mo, S. Yan, Concrete early-age strength monitoring using embedded piezoelectric transducers, Smart Mater. Struct. 15 (2006) 1837.

[28] H.-J. Jung, I.-H. Kim, S.-J. Jang, An energy harvesting system using the wind-induced vibration of a stay cable for powering a wireless sensor node, Smart Mater. Struct. 20 (2011), 075001.

[29] S. Ali, M. Friswell, S. Adhikari, Analysis of energy harvesters for highway bridges, J. Intell. Mater. Syst. Struct. 22 (2011) 1929–1938.

[30] M. Peigney, D. Siegert, Piezoelectric energy harvesting from traffic-induced bridge vibrations, Smart Mater. Struct. 22 (2013).

[31] C.K. Soh, K.K.H. Tseng, S. Bhalla, A. Gupta, Performance of smart piezoceramic patches in health monitoring of a RC bridge, Smart Mater. Struct. 9 (2000) 533–542.

[32] F. Duarte, A. Ferreira, P. Fael, Road pavement energy–harvesting device to convert vehicles' mechanical energy into electrical energy, J. Energy Eng. 144 (2018).

[33] X. Xu, D. Cao, H. Yang, M. He, Application of piezoelectric transducer in energy harvesting in pavement, Int. J. Pavement Res. Technol. 11 (2018) 388–395.

[34] Z. Zhang, H. Xiang, Z. Shi, Modeling on piezoelectric energy harvesting from pavements under traffic loads, J. Intell. Mater. Syst. Struct. 27 (2016) 567–578.

[35] H. Zhao, J. Ling, J. Yu, A comparative analysis of piezoelectric transducers for harvesting energy from asphalt pavement, J. Ceram. Soc. Jpn. 120 (2012) 317–323.

[36] N. Kumari, A. Trivedi, Application of semi-active control strategy for the wall retaining granular fills, in: Proceedings of China-Europe Conference on Geotechnical Engineering, Springer, 2018, pp. 978–982.

[37] N. Kumari, A. Trivedi, Semi-active control strategy for horizontal dynamic loading on wall retaining granular fills. sustainable civil engineering practices, in: Lecture Notes in Civil Engineering, 72, Springer, Singapore, 2020.

[38] H.W. Kim, A. Batra, S. Priya, K. Uchino, D. Markley, R.E. Newnham, H. F. Hofmann, Energy harvesting using a piezoelectric "cymbal" transducer in dynamic environment, Jpn. J. Appl. Phys. 43 (2004) 6178.

[39] H. Zhao, J. Yu, J. Ling, Finite element analysis of Cymbal piezoelectric transducers for harvesting energy from asphalt pavement, J. Ceram. Soc. Jpn. 118 (2010) 909–915.

[40] J. Wang, Z. Shi, H. Xiang, G. Song, Modeling on energy harvesting from a railway system using piezoelectric transducers, Smart Mater. Struct. 24 (2015), 105017.

[41] S. Zhao, A. Erturk, Deterministic and band-limited stochastic energy harvesting from uniaxial excitation of a multilayer piezoelectric stack, Sens. Actuators A: Phys. 214 (2014) 58–65.

[42] Z. Zhang, L. Tang, H. Xiang, Piezoelectric energy harvesting from bridge vibrations using different models for moving vehicles, J. Aerospace Eng. 32 (2019).

[43] P. Negi, R. Chhabra, N. Kaur, S. Bhalla, Health monitoring of reinforced concrete structures under impact using multiple piezo-based configurations, Constr. Build. Mater. 222 (2019) 371–389.

[44] Z. Zhang, H. Xiang, Z. Shi, J. Zhan, Experimental investigation on piezoelectric energy harvesting from vehicle-bridge coupling vibration, Energy Convers. Manage. 163 (2018) 169–179.

[45] S.-H. Kim, J.-H. Ahn, H.-M. Chung, H.-W. Kang, Analysis of piezoelectric effects on various loading conditions for energy harvesting in a bridge system, Sens. Actuators A: Phys. 167 (2011) 468–483.

[46] Y. Zhang, S.C. Cai, L. Deng, Piezoelectric-based energy harvesting in bridge systems, J. Intell. Mater. Syst. Struct. 25 (2014) 1414–1428.

[47] A. Erturk, Piezoelectric energy harvesting for civil infrastructure system applications: Moving loads and surface strain fluctuations, J. Intell. Mater. Syst. Struct. 22 (2011) 1959–1973.

[48] A. Erturk, D.J. Inman, An experimentally validated bimorph cantilever model for piezoelectric energy harvesting from base excitations, Smart Mater. Struct. 18 (2009).

[49] A. Erturk, D.J. Inman, On mechanical modeling of cantilevered piezoelectric vibration energy harvesters, J. Intell. Mater. Syst. Struct. 19 (2008) 1311–1325.

[50] S. Roundy, P.K. Wright, J. Rabaey, A study of low level vibrations as a power source for wireless sensor nodes, Computer Commun. 26 (2003) 1131–1144.

[51] J.A. Paradiso, T. Starner, Energy scavenging for mobile and wireless electronics, IEEE Pervas. Comput. 4 (2005) 18–27.

[52] L. Guo, Q. Lu, Modeling a new energy harvesting pavement system with experimental verification, Appl. Energy. 208 (2017) 1071–1082.

[53] Y. Song, Finite-element implementation of piezoelectric energy harvesting system from vibrations of railway bridge, J. Energy Eng. 145 (2019), 04018076.

[54] S. Jing-song, H. XIAO-MING, L. Gong-yun, Dynamic response analysis of pavement structure under moving load, J. Highw. Transp. Res. Dev. 1 (2007).
[55] B. Patzák, OOFEM—an object-oriented simulation tool for advanced modeling of materials and structures, Acta Polytech. 52 (2012).
[56] K.-q. Qi, Y. Xiang, C. Fang, Y. Zhang, C.-s. Yu, Analysis of the displacement amplification ratio of bridge-type mechanism, Mech. Mach. Theory. 87 (2015) 45–56.
[57] Y. Amini, M. Heshmati, P. Fatehi, S.E. Habibi, Energy harvesting from vibrations of a functionally graded beam due to moving loads and moving masses, J. Eng. Mech. 143 (2017).
[58] P. Bonello, S. Rafique, Modeling and analysis of piezoelectric energy harvesting beams using the dynamic stiffness and analytical modal analysis methods, J. Vib. Acoust. 133 (2011).
[59] N.T. Khiem, T.T. Hai, L.Q. Huong, Effect of piezoelectric patch on natural frequencies of Timoshenko beam made of functionally graded material, Mater. Res. Express. 7 (2020), 055704.
[60] H. Roshani, S. Dessouky, A. Montoya, A. Papagiannakis, Energy harvesting from asphalt pavement roadways vehicle-induced stresses: a feasibility study, Appl. Energy. 182 (2016) 210–218.
[61] A. Jasim, H. Wang, G. Yesner, A. Safari, A. Maher, Optimized design of layered bridge transducer for piezoelectric energy harvesting from roadway, Energy 141 (2017) 1133–1145.
[62] A.F. Jasim, H. Wang, G. Yesner, A. Safari, P. Szary, Performance analysis of piezoelectric energy harvesting in pavement: laboratory testing and field simulation, Transp. Res. Rec. 2673 (2019) 115–124.
[63] J. Zhao, H. Wang, Mechanistic modeling and economic analysis of piezoelectric energy harvesting potential in airport pavements, Transp. Res. Rec. 2674 (2020) 64–75.
[64] H. Roshani, P. Jagtap, S. Dessouky, A. Montoya, A. Papagiannakis, Theoretical and experimental evaluation of two roadway piezoelectric-based energy harvesting prototypes, J. Mater. Civ. Eng. 30 (2018), 04017264.
[65] C. Wang, J. Zhao, Q. Li, Y. Li, Optimization design and experimental investigation of piezoelectric energy harvesting devices for pavement, Appl. Energy. 229 (2018) 18–30.
[66] X. Jiang, Y. Li, J. Li, J. Wang, J. Yao, Piezoelectric energy harvesting from traffic-induced pavement vibrations, J. Renew. Sustain. Energy 6 (2014), 043110.
[67] N. Kumari, A. Trivedi, Factors influencing piezoelectric response of horizontally and vertically embedded PZT patch in confined granular fills, Adv. Mater. Sci. Eng. 2022 (2022).
[68] L. Gammaitoni, I. Neri, H. Vocca, Nonlinear oscillators for vibration energy harvesting, Appl. Phys. Lett. 94 (2009), 164102.
[69] B. Mann, B. Owens, Investigations of a nonlinear energy harvester with a bistable potential well, J. Sound Vibrat. 329 (2010) 1215–1226.
[70] L. Cveticanin, M. Zukovic, D. Cveticanin, Non-ideal source and energy harvesting, Acta Mech 228 (2017) 3369–3379.
[71] P.L. Green, E. Papatheou, N.D. Sims, Energy harvesting from human motion and bridge vibrations: an evaluation of current nonlinear energy harvesting solutions, J. Intell. Mater. Syst. Struct. 24 (2013) 1494–1505.
[72] P. Cahill, A. Mathewson, V. Pakrashi, Experimental validation of piezoelectric energy-harvesting device for built infrastructure applications, J. Bridge Eng. 23 (2018).
[73] Y. Zhang, C. Cai, W. Zhang, Experimental study of a multi-impact energy harvester under low frequency excitations, Smart Mater. Struct. 23 (2014), 055002.
[74] A. Erturk, D.J. Inman, An experimentally validated bimorph cantilever model for piezoelectric energy harvesting from base excitations, Smart Mater. Struct. 18 (2009), 025009.
[75] M. Rhimi, N. Lajnef, Tunable energy harvesting from ambient vibrations in civil structures, J. Energy Eng. 138 (2012) 185–193.
[76] M. Wischke, M. Masur, M. Kröner, P. Woias, Vibration harvesting in traffic tunnels to power wireless sensor nodes, Smart Mater. Struct. 20 (2011), 085014.
[77] P. Cahill, N.A.N. Nuallain, N. Jackson, A. Mathewson, R. Karoumi, V. Pakrashi, Energy harvesting from train-induced response in bridges, J. Bridge Eng. 19 (2014), 04014034.
[78] Y. Song, Finite-Element implementation of piezoelectric energy harvesting system from vibrations of railway bridge, J. Energy Eng. 145 (2019).
[79] B. Shen, X. Yang, Z. Li, A cement-based piezoelectric sensor for civil engineering structure, Mater. Struct. 39 (2006) 37–42.
[80] M. Sun, Z. Li, X. Song, Piezoelectric effect of hardened cement paste, Cem. Concr. Compos. 26 (2004) 717–720.
[81] H. Rong, S. Zhifei, Exact analysis of 0-3 cement-based piezoelectric composites, J. Intell. Mater. Syst. Struct. 22 (2011) 221–229.
[82] E. Ghafari, Y. Yuan, C. Wu, T. Nantung, N. Lu, Evaluation the compressive strength of the cement paste blended with supplementary cementitious materials using a piezoelectric-based sensor, Constr. Build. Mater. 171 (2018) 504–510.
[83] R. Tawie, H.-K. Lee, Monitoring the strength development in concrete by EMI sensing technique, Constr. Build. Mater. 24 (2010) 1746–1753.
[84] H. Gong, Z. Li, Y. Zhang, R. Fan, Piezoelectric and dielectric behavior of 0-3 cement-based composites mixed with carbon black, J. Eur. Ceram. Soc. 29 (2009) 2013–2019.
[85] J.W. Yang, H.P. Zhu, J. Yu, D.S. Wang, Experimental study on monitoring steel beam local corrosion based on EMI technique, Appl. Mech. Mater. Trans Tech Publ. 273 (2013) 623–627.
[86] A. Dwivedi, A. Banerjee, B. Bhattacharya, Simultaneous energy harvesting and vibration attenuation in piezo-embedded negative stiffness metamaterial, J. Intell. Mater. Syst. Struct. 31 (2020) 1076–1090.
[87] Z. Yu, A. Zhou, D. Lau, Mesoscopic packing of disk-like building blocks in calcium silicate hydrate, Sci. Rep. 6 (2016) 1–8.
[88] Z. Li, D. Zhang, K. Wu, Cement-based 0-3 piezoelectric composites, J. Am. Ceram. Soc. 85 (2002) 305–313.
[89] B. Dong, Z. Li, Cement-based piezoelectric ceramic smart composites, Compos. Sci. Technol. 65 (2005) 1363–1371.
[90] S. Wen, D. Chung, Cement-based materials for stress sensing by dielectric measurement, Cem. Concr. Res. 32 (2002) 1429–1433.
[91] H. Gong, Y. Zhang, J. Quan, S. Che, Preparation and properties of cement based piezoelectric composites modified by CNTs, Curr. Appl. Phys. 11 (2011) 653–656.
[92] Z. Shi, J. Wang, Dynamic analysis of 2-2 cement–based piezoelectric transducers, J. Intell. Mater. Syst. Struct. 24 (2013) 99–107.
[93] J.T. Yao, Concept of structural control, Journal of the Structural Division 98 (1972) 1567–1574.
[94] Z. Li, H. Gong, Y. Zhang, Fabrication and piezoelectricity of 0–3 cement based composite with nano-PZT powder, Curr. Appl. Phys. 9 (2009) 588–591.
[95] F. Wang, H. Wang, Y. Song, H. Sun, High piezoelectricity 0–3 cement-based piezoelectric composites, Mater. Lett. 76 (2012) 208–210.
[96] A. Chaipanich, N. Jaitanong, Effect of poling time on piezoelectric properties of 0-3 PZT-portland cement composites, Ferroelectr. Lett. 35 (2008) 73–78.
[97] K.H. Lam, H. Chan, Piezoelectric cement-based 1-3 composites, Appl. Phys. A. 81 (2005) 1451–1454.
[98] N. Jaitanong, R. Yimnirun, H. Zeng, G. Li, Q. Yin, A. Chaipanich, Piezoelectric properties of cement based/PVDF/PZT composites, Mater. Lett. 130 (2014) 146–149.
[99] A. Chaipanich, Dielectric and piezoelectric properties of PZT–silica fume cement composites, Curr. Appl. Phys. 7 (2007) 532–536.
[100] S. Banerjee, K.A. Cook-Chennault, Influence of aluminium inclusions on dielectric properties of three-phase PZT–cement–aluminium composites, Adv. Cem. Res. 26 (2014) 63–76.
[101] R. Rianyoi, R. Potong, A. Ngamjarurojana, A. Chaipanich, Microstructure and electrical properties of 0-3 connectivity barium titanate− Portland cement composite with 40% barium titanate content, Ferroelectr. Lett. Sec. 43 (2016) 59–64.
[102] Roshani H., Dessouky S., Papagiannakis A.: Experimental investigation of energy harvesting prototypes for asphalt roadways. pp 10-19. (2017).



Nisha Kumari is pursuing her doctoral research at the Department of Civil Engineering, Delhi Technological University Delhi, India. She obtained her B.Tech. (Honours) in civil engineering from University College of Engineering, Rajasthan Technical University Kota, India, in 2012 and M.Tech. in Soil mechanics and foundation engineering from National Institute of Technology Kurukshetra, India, in 2014. She does research in Geotechnical engineering. Her current project is "Dynamic Response Analysis of Wall Retaining Granular Fills." She likes Yoga, cycling, traveling, playing tennis and reading self-improvement books.



A. Trivedi served as a Professor of Civil Engineering and Dean of Industrial Research and Development at Delhi Technological University, Delhi, India. His previous affiliations include Professor of Civil Engineering and Dean, Faculty of Technology at the University of Delhi, Dean of Continuing Education at Delhi College of Engineering, Head of the Department of Civil Engineering at the University of Delhi, Head of the Department of Civil Engineering at Delhi Technological University, Delhi, India. He obtained his Bachelor's and Master's degrees from REC (now NIT), Kurukshetra and Ph.D. from TIET (Thapar Institute of Engineering and Technology), Patiala, India. He has several research papers as lead author in reputed journals, conferences and symposia. He is a fellow of the Indian Association of Structural Engineers and an active member of several professional societies namely, ASCE, ISRMTT, ISTE, and IGS, Delhi. He has supervised many prestigious consultancy projects. He has guided more than fifty Master's and Ph.D. candidates to award degrees at Thapar University, Patiala, the University of Delhi and Delhi Technological University, Delhi, India. He has a couple of PCTs to his credit. His current interests include mechanics of fractured masses, hydraulic flow amid cemented and un-cemented ash fills, liquefaction of soils, and bio-cemented soils and rocks. He likes playing tennis and participates in Yoga and ontological studies community activities.

2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT) | 978-1-6654-1005-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICICICT54557.2022.9917632

A Survey on Integration of Blockchain and IoT (BIoT): Open Issues, Challenges & Solution

Chirag Garg[1], Deepak Kumar Mishra[2],Dikshant Raj[3],Pawan Singh Mehra[4]
[1,2,3,4]Department of Computer Science and Engineering, Delhi Technological University, New Delhi, India
[1]cgarg4923@gmail.com, [2]deepakmishra.it@gmail.com, [3]dikshantraj09@gmail.com, [4]pawansinghmehra@gmail.com

Abstract—Blockchain and IoT are becoming the buzzwords in the tech world. Blockchain in itself is a very vast topic with various applications and uses but it has its fair share of challenges. Since IoT consists of a vast number of devices, from batterypowered devices with less computational resources to powerful machines, connected over a network it also has hurdles with various barriers. But this paper provides an in-depth review of the current situation of integration of Blockchain and the Internet of Things(IoT) and the Security Issues or Challenges that arise. It starts by discussing the ins and outs of blockchain technology and IoT and their applications. It compares the challenges that arise from using different public blockchain technologies like scalability, interoperability, computational resource constraints, storage, etc. This review also analyzes the current solutions that are present to mitigate the above-mentioned challenges in the integration of Blockchain and IoT. This paper also shines some light on the applications of BIoT and how it is changing and influencing the security challenges in current IoT devices. A detailed and in-depth analysis of this integration of blockchain and IoT was missing so this paper tries to summarize the challenges and solutions.

Index Terms—Blockchain, IoT, BIoT, CIA, Challenges, Merkle, Decentralized

I. Introduction

With the ever-evolving world, the number of devices associated with the Internet of Things(IoT) is increasing exponentially. IoT has laid its emphasis on various aspects from the supply chain to smart cities, smart homes to smart healthcare, and many more. The Internet Of Things (IoT) is associated with the internet of physical devices that we call "Things". IoT is an emerging network that joins as myriad things as possible, like RFID, global positioning systems(GPS), infrared sensors, laser scanners. The system joins all available devices to the Internet for the exchange of information and communication, in order to locate, track, identify, monitor, and manage the whole network intelligently [29] - [32]. Due to an innumerable amount of devices in IoT, The IoT demands low latency for the sake of speed, and elevated throughput to support this evergrowing number of devices, in a wired or wireless fashion. One factor that undoubtedly makes IoT unique is the heterogeneity of the devices involved with distinctive functionalities in the human as well as machine centric environment to cope with the future. The unparalled data acquisition, exchange among diverse peer devices initializes the integration of the cyber in the real-world and enhances decision-building processes. Nonetheless, such features of IoT also direct us to the security concerns related to the devices associated. Security is of utter importance these days. Even the tiniest point of vulnerability can cause serious damage, be it in terms of a privacy breach or unauthorized access to the devices. Security of any device or person over the internet is narrated in terms of three principle aspects popularly known as the "CIA" triad. The CIA is short for Confidentiality, Integrity, and Availability. Confidentiality touch on the art of keeping the data private from those who are not authorized access to that particular information [33]. Since millions of IoT devices share enormous data every second it is very crucial to maintain the confidentiality of data. Integrity refers to the fact that the data sent is either complete or whole. In simpler terms, the data has not been vandalized and is completely reliable and authentic. The concept of Non-Repudiation which is the inability to deny something perfectly aligns with the Integrity constraint. Availability on the other side signifies that the resource is accessible when the user demands it. It ensures that the authorized users have reliable access to the resource at any time. For instance, a home equipped with smart appliances connected through the IoT technology can be intruded on by an unauthorized party who may switch states of the appliances frequently leading to the burning down of the circuit of devices which may lead to fire hazards. Another example can be the automated supply chain of Ice Cream, If an intruder gained access to the sensors installed in the supply truck, it may lead to false data being sent to the server causing a gigantic loss to the company [12]. So security in IoT devices is of utmost importance. The traditional IoT architectures lay emphasis on a more centralized controller, making the entire IoT traffic channel through a single platform like a server or a distributed architecture with some integration of centralized control. But this centralized control of the IoT architecture may contribute to a single point of failure and also if any unauthorized entity gained access to this centralized controller, then he/she might gain the control of the entire network altogether. However, with the growth of device count in the IoT network and rapidly changing technology, the IoT architecture is making a transition to the decentralized architecture which targets various aspects all at once such as the higher cost of construction and maintenance of large-scale data centers, performance bottlenecks, and more security and privacy risks [3]. When we talk about decentralized technology, Blockchain is the one-stop solution. The surfacing of blockchain technology has opened many doors that target a wide variety of threats to IoT networks. Blockchain is a peer-to-peer network

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:39:15 UTC from IEEE Xplore. Restrictions apply.

or it is a distributed ledger that does not belong to a specific individual but rather is owned by each entity present in the network [6]. Blockchain is an immutable and decentralized ledger. One such example is Bitcoin which uses Blockchain technology. In 2008, S. Nakamoto introduced the theory of a blockchain which gained enormous attention over the recent years as a poping peer-to-peer (P2P) technology for distributed computing and decentralized data sharing of digital assets. A bitcoin is purely peer-to-peer electronic cash which allows this cash to be directly transferred between the interacting parties without the intervention of any third centralized party like a bank [11]. A blockchain is just like a linked list in which each block or node depends on its own data along with the hash of the previous block. A node in a blockchain network consists of a previous hash, nonce, timestamp, and Merkle root. A Merkle tree or Merkle hash tree is a way to generate a cryptographic hash tree which is one way only [34]. Then using the authentication path information every leaf node is verified which decreases the cost of verification as only the hash needs to be calculated [13]. This means that a new transaction or a node is added to the existing blockchain only if it is validated by all the peers present in the network which implies that each peer will have the information of the alterations in the blockchain and once added it cannot be altered. Since Blockchain is decentralized in nature, it encourages a secure and encrypted network for IoT nodes and thus eliminates a single point of weakness making it strenuous for intruders to control the entire network. Hence it not only targets the scalability, trust cost, and singular failure but also promotes smart and efficient transactions. Thus blockchain when applied to an IoT network provides security, privacy, and transparency and also removes the trust factor among the communicating devices. Blockchain enables security in a network using the consensus algorithm. Consensus refers to the concept a new node is only added to the network if all the other existing nodes agree upon it. There are various consensus algorithms.

- PoW - Proof of Work
- PoS - Proof of Stake
- PBFT - Practical Byzantine Fault Tolerance
- PoB - Proof of Burn
- PoTE - Proof of Time Elapsed
- PoC - Proof of Capacity

Lastly, integrating IoT with Blockchain will strengthen the self-decision and trust issues among the communicating IoT devices extending the security it provides and it thus is making a transition from the traditional client-server architecture. The remainder of this article is composed as follows. Section II talks about the brief background and related work in integrating blockchain and IoT along with security aspects. Section III gives a glimpse of the challenges in the field of IoT, Blockchain, and their integration. Section IV provides details about the existing solution to this integration problem and Section V concludes this.

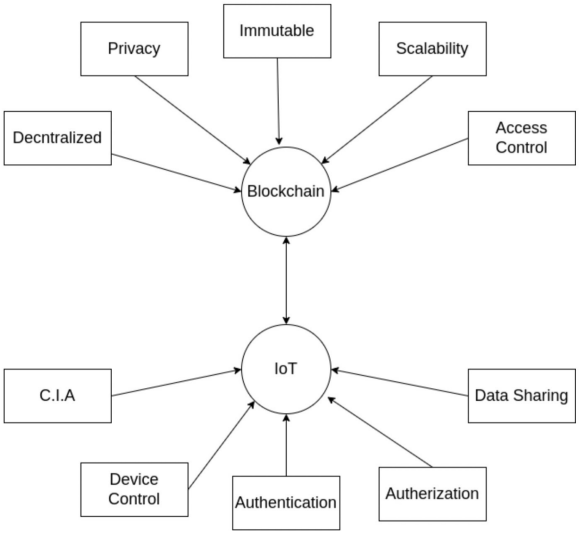


Fig. 1. Issues in Blockchain and IoT

II. Related Work

The convergence of Blockchain, IoT, and Its security has been published in various surveys and papers. An exhaustive survey of the existing blockchain (Public) for IoT devices especially resource-constrained devices for a review of all the possible security vulnerabilities on a BIoT (Blockchain+IoT) network was done [1]. An extensive survey of distinct varieties of blockchains was performed in order to preserve the security of resource-constrained IoT devices. Which provided a basis for a study that encloses the issues and challenges faced by the IoT integration with Blockchain devices. This paper throws some light on available resource-constrained IoT devices along with the protocols used for communication and their classification. A comprehensive analysis of the challenges needed to be overcome for successful integration of Blockchain and IoT in resource-constrained IoT devices. Proposed solutions and their feasibility was also analyzed [1]. Different types of blockchains were summarized. Then, an overview of the applications in the domain of blockchain along with IoT technology was provided in this paper. Taxonomy and a juxtaposition between the various state art methods for securing and preserving the privacy of blockchain technologies according to the blockchain model were done with certain security limitations, goals, complexity of computation, and communication overhead [2]. Integration of blockchain with IoT is named BCoT. This paper provides an exhaustive survey on Blockchain+IoT (BCoT). In particular, The paper then discussed the opportunities of Blockchain+IoT and shows the structure of Blockchain+IoT. We further shape the research problems in blockchain for the upcoming generation networks. We further discuss the potential applications of Blockchain+IoT and shape the open research directions in Blockchain+IoT [3]. Security concerns along with their corresponding solutions, challenges in research, as well as the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:39:15 UTC from IEEE Xplore. Restrictions apply.

future scope and directions in the integration of smart contracts with IoT for the first time, with an emphasis on three crucial domains:

1 Particularities that are inherently vulnerable.
2 Vulnerabilities in programming. 3 feasibility attacks.

It performs an exclusive and in-depth review of well known security challenges and potential research openings associated with the deploying of smart contracts in an IoT setting [4]. A Systematical analysis of the state of the art of IoT security particularly based on the blockchain, devoting special attention to the present security features, approaches, technologies, issues and connected plots in blockchain implanted with IoT. The implantation along with the interoperation of blockchain and IoT is an crucial development in the communication system of computation. Blockchain can generate data protection, access authentication, and anti-DDoS concern on a huge number of devices, anonymity protection, and granting privacy for IoT devices, and decreasing one points through ductile data interaction and consensus of nodes. Inconsistency in block generation happens due to the huge difference in the computing and capabilities of communication of IoT devices. Furthermore, accessing enormous devices will result in a huge peak of data in the blockchain [5]. IoT manages security, secrecy, and accessibility which depend on conventional cryptographic arrangements [6].

Wireless Sensor Network, also known as WSN, is a heart component for running IoT systems. WSN and its challenges, most importantly in terms of security perspective through exposing the vulnerabilities in WSN. Designing an absolutely new WSN Security system with blockchain integration was performed[7]. Applications of blockchain in the latest technology with enhanced execution (reasonable idleness, increased throughput), appropriate low exchange expense, and pliable viability were used to judge the challenges [8]. Specific security mechanisms are adopted by the popular communication protocols for maintaining security [9]. Highlighting the security along with privacy issues in distinct fields of Bitcoins and cryptocurrency major components of Bitcoin, its significant characteristics, and their related concepts were pointed out in brief for enhanced understanding [10]. Blockchain is an immutable type of ledger that has the capability of storing data. It has its applications in the smart grids for cybersecurity. A variety of possible security attacks are possible in smart grids to which blockchain can provide an efficient solution [33]

TABLE I
RELATEDWORKON BIOT

Year	Author	Main Focus and Contribution
2021	Jing Huey Khor et al.	Survey public blockchains for the security of resource-constrained IoT devices.
2019	Mohamed Amine Ferrag et al.	Application domains of blockchain technologies in IoT
2019	Hong-Ning Dai et al.	survey of BCoT interoperability, privacy, security, and traceability and reliability
2021	Kai Peng et al.	the security issues, corresponding solutions, research challenges, as well as future directions in the integration between smart contracts and the IoT
2021	Li Da Xu et al.	Analyzes the state of the art of IoT security based on the blockchain, paying special attention to the security features, issues, technologies, approaches, and related scenarios in blockchain-embedded IoT. The integration and interoperation of blockchain and IoT
2019	Devibala A.	blockchain and delineating the present examinations that arrangement with blockchain progressions.
2022	Swathi Darla, Naveena C, et al.	WSN and its challenges, most importantly in terms of security perspective
2022	Sougata Biswas, et al.	Applications of blockchain in latest technology
2019	Francesca Meneghello, et al.	security vulnerabilities in Real IoT devices
2018	Mauro Conti, et al.	Highlighting the security and privacy issues in different fields of Bitcoin.
2021	Peng Zhuang, et al.	Development of smart grids to make them more resilient against cyberattacks

III. OPEN ISSUES AND CHALLENGES IN INTEGRATING BLOCKCHAIN & IOT

There are various challenges associated with public blockchain which need to be summoned before embedding it in the IoT ecosystem. These challenges are heterogeneous in nature but majorly they are more or less related to the size of the ledger, scalability, and last but not least resources required for computation of the consensus protocol [1]. Moreover, Resource-Constrained IoT Devices, as the name suggests have very limited CPU utilities and small storage capacities so that these devices could perform at maximum efficiency at the expense of minimal energy/power. Due to their limited storage and low computational power, these IoT devices are incapable of storing substantial ledger sizes and high computational resources. The low-transaction rates of the existing blockchain pose additional hurdles inhibiting the blockchain from incorporating more and more IoT devices. Blockchain although being a highly secure and trustworthy tool still faces some challenges:

A. Challenges in Blockchain

- Leakage of Private Key: ECDSA, which is short for elliptical curve digital signature algorithm, used to generate a private key for the user is highly vulnerable in nature [24]. If limited randomness is deployed to generate a private key, then the intruder can easily recover the private key and can exploit the data in numerous ways such as a replay attack, tampering attack, key attack, and impersonation attack, modification attack, and man-inthe-middle attack.
- 51% attack: Consensus algorithms are one of the most important aspects of the blockchain technology as a new node/transaction will only be added to the existing

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:39:15 UTC from IEEE Xplore. Restrictions apply.

blockchain if all the other nodes verify it. If a miner is enjoying more than 50% of the total computational power of the blockchain, then the blockchain is prone to a 51% attack. Under a 51% attack, the blockchain is susceptible to the following attacks.

  - Information Tampering
  - Reordering and exclusion of transaction
  - Hamper normal mining operations
  - Interfere with transaction confirmation
  - Execute double-spending and initialize a reverse transaction [2]

- Double Spending: Instead of transfer of the actual asset, a copy of the asset being sent to the receiving party is termed double-spending. For instance, If you want to send an image to another person using email, you will simply go to Gmail or any other mail service and type the recipient's email address, then select the image and simply send it. Now, instead of the actual image being sent to the recipient, a copy of the image will be shared with that person. So a copy of the actual asset is being transferred rather than the actual asset. This vulnerability can be easily used by the intruder in the following way:
  - An intruder transmits the same coin in prompt successions to 2 distinct addresses and if the buyer accepts the coin without waiting for the confirmation, then the buyer is caught in the web of doublespending as he/she will not be able to get a receipt of the transaction.

B. Challenges in IoT

- Heterogeneity of Devices: Heterogeneity in IoT devices is a major concern when it comes to communication and data sharing among IoT devices. Heterogeneity may be due to heterogeneous communication protocols or due to heterogeneous data types in terms of structured, unstructured, or semi-structured data. Heterogeneity may give rise to further issues like interoperability, privacy, etc. [3]
- Interoperability: It refers to the ability or efficiency of IoT systems to make use of information by exchanging info and collaborating with each other. IoT systems can be both software and hardware. It becomes extremely difficult to exchange info between Industrial units, strategic centers, and IoT systems due to the decentralization and heterogeneity of IoT networks.
- Resource Constraint: IoT devices namely sensors, RFID tags, actuators, and smart meters due to limited resources availability lack computing resources, storage resources, and battery power and hence are unable to carry out complex computations such as consensus. For instance, Passive RFID tags can only draw energy from RFID readers or from a suitable environment due to the absence of battery power [25]. Resource-Constrained devices are very much susceptible to malicious attacks due to their limited resource availability.

C. Challenges in Integrating Blockchain with IoT

- Scalability: Transactional Latency and Block Latency are the two major things that affect the scalability of a blockchain significantly. Whenever a new block or a transaction is added/validated in a blockchain it consumes some time due to the consensus mechanism as the block or the transaction needs to be validated first, So this delay in the addition of a new block/transaction is known as Block/Transaction Latency. The smart city application in the European Union has projected the number of IoT devices to touch 53.63 Million by 2025 [26]. As the number of nodes in a blockchain grows, the transmission overheads also increase posing a threat to real-time transmission. In terms of BIoT(Blockchain+IoT) in the food supply chain, Real-time transmission of data is a must, as it is crucial to measure the humidity, temperature, etc of the food every instant instantaneously [12].
- Computational Resources: There are a lot of consensus algorithms available out there, some of which are PoW (Proof of Work) used by bitcoin, PoS (Proof of Stake), and many more algorithms. Out of these algorithms, PoW demands the highest computational power. Other consensus algorithms despite not requiring heavy computational like PoW but still need some computational power to validate blocks and transactions. For instance, Algorand Blockchain, a node of which uses pure Proof of Stake(PoS) consensus algorithm prerequisites a system to exhibit 4 to 8 Gigabytes of RAM, a 32/64 bit 4 core CPU and a stable 100 Megabits per second of internet speed connection [28]. Thus ,intensive and complex computations to be performed by node inhibits resource-constrained devices to directly participate in the consensus protocol.
- Storage Size: There are various types of nodes in a blockchain two of which are a full node and a lightweight node. A full node is a node that contains the entire blockchain in order to validate a new block/transaction due to which its size is enormous. A lightweight node on the other hand contains only the headers of the block instead of the whole copy. A lightweight node uses SPV (Simple Payment Verification) to validate the payment [27]. Each lightweight node must be attached to a full node as it helps lightweight to transmit a transaction to the blockchain network. The IoT devices associated with BIoT (Blockchain+IoT) are prescribed to store this blockchain ledger at IoT nodes. The current size of Bitcoin Blockchain is around 412 GB and that of Etherium full archive node is around 9TB (up to November 2021). The storage capacity of Resource-constrained IoT devices is generally 216KB and hence it is incapable of storing both full and lightweight nodes making it difficult to integrate with blockchain [1].

IV. Existing Solutions to Integration challenges

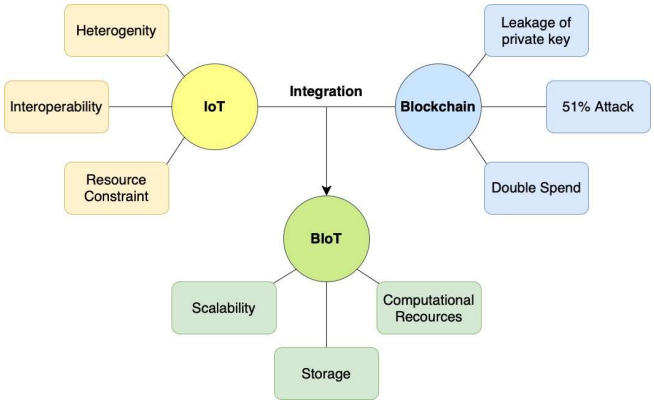
Some of the challenges related to the integration of Blockchain with IoT are storage, computational power, scal-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:39:15 UTC from IEEE Xplore. Restrictions apply.

Fig. 2. Challenges in Integration of Blockchain and IoT

ability, and efficiency of energy [14]. These are some of the shortcomings that need to be addressed in order to integrate blockchain with IoT (BIoT). Some of the solutions that tackle these issues are:

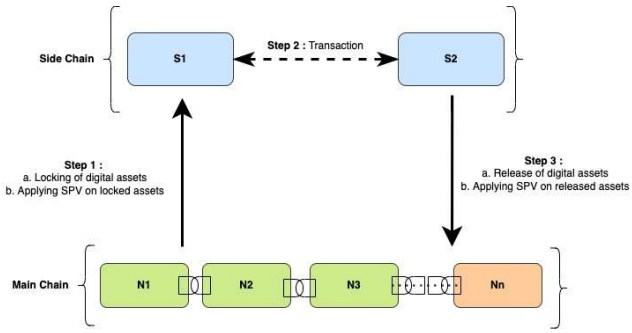
- Scalability: As the number of nodes increases in a blockchain network, the time taken to validate a transaction increases, So it becomes difficult to scale up the IoT network. In order to improve scalability the following useful measures are:
  - Reducing Latency: Zilliqa, a public blockchain, enhanced the classical Practical Byzantine Fault Tolerance (PBFT) by using multi-signature and hence reducing the computational overheads [15]. SPoS, Secure Proof of Stake, is another refined version of the PBFT consensus algorithm that is used by Elrond Blockchain. SPoS is a mixture of both SPoS and PBFT consensus algorithms [16]. The enhanced PBFT consensus algorithm has a remarkably low latency when compared to the PoW consensus used by Bitcoin which is nearly 600 s. Enhances PBFT used by Zilliqa has a latency of 120 s while that of SPoS of Elrond is about 4 s [17]. Hence by reducing the latency, we can easily scale our IoT network and can promote real-time data transfer in the IoT network.



  - Side Chaining: A sidechain is nothing but a chain that runs parallel to the main chain and cooperate with the main chain by utilizing a two-way peg mechanism. This architecture reduces the load on the main chain by performing computations on the side chain and exchanging digital assets between these chains. In this approach, the tokens are locked at the main chain allowing the transactions to proceed at the side chain or vice versa [18]. This side chaining mechanism uses Simplified Payment Verification(SPV) protocol to validate transactions between both the chains. This concept of side chaining increases the throughput significantly and reduces the computational overhead allowing IoT networks

Fig. 3. Symmetric two way peg mechanism side chain

  - Sharding: The procedure of assembling blockchain nodes into minuscule groups called shards is termed sharding. There are various ways to group nodes into smaller chunks such as based on addresses. We can club all the nodes having two consecutive zeroes at the beginning of the address in one shard. Similarly, we can put all the nodes having addresses starting with 01 in one shard. Sharding can significantly improve scalability by permitting parallel transactions to be validated at once. But this method suffers from single-shard takeovers. In this attack, the attacker takes over a shard by capturing multiple/majority nodes in a shard. Ethereum, a public blockchain, uses sharding, and to prevent this attack Ethereum randomly assigns a new node to a shard to be validated. Cross-sharding communication protocol is used by blockchain networks to allow communication between these shards. However, security measures must be designed carefully for these intra-shard communications.
- Storage Size: Due to the limited storage and limited CPU of the resource-constrained IoT devices, it becomes extremely burdensome to carry out large computations on these devices. Some of the cures for these problems are:
  - Off-chain: An innovative solution to tackle this to scale rapidly. See Fig 3 for reference.



storage issue is using off-chain which basically stores the hash of data related to the transaction instead of storing the complete nodes of the blockchain. The hash will then serve as the key to access the whole data stored off the chain, off-chain storage. The decentralized nature of the off-chain storage can be used by the blockchain network by delivering access to the existing data inside the peer-to-peer network. Filecoin, Interplanetary File System(IPFS), Swarm, and many more are some examples of decentralized file storage systems. The efficiency of this approach is not significant as the user needs to maintain three distinct software in order to maintain and utilize the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:39:15 UTC from IEEE Xplore. Restrictions apply.

main chain, the decentralized storage, and the middleware. To handle these three software simultaneously, multiChain platforms appear to be a one-stop solution by integrating all these software into a single platform [1].

- Pruning: It refers to the process of removing redundant or non-required information to reduce the local size of the blockchain which helps in increasing the efficiency of transactions. In the Bitcoin blockchain, the pruning mechanism allows the user to store the last 550 nodes instead of storing the complete blockchain for validation of a transaction [19]. However, for the pruning process to take place one needs to download the entire blockchain first which contributes to the drawback of this solution. Monero, a private anonymous blockchain, has developed a pruning function to prune the blockchain size by two-thirds of its original size. For instance, if the Monero blockchain size is 90GB, then after pruning only 30GB has to be stored [20]. Snapshot is also an interesting solution to this problem. It basically clubs several transactions together to the same address into a single record. A global snapshot used by the IOTA ledger in order to decrease ledger size is one such example [21]. But this approach suffers from data loss which can be easily solved using Permanode, a node that keeps whole data permanently.

- Computational Resources: Due to the limited storage and limited CPU of the resource-constrained IoT devices, it becomes extremely burdensome to carry out large computations on these devices. Some of the cures for these problems are:
  - Cloud Offload: Resource-constrained IoT devices suffer from very low computational power due to their limited CPU processing. Hence it becomes extremely difficult for these devices to perform heavy computations on themselves. So in order to solve this problem Cloud servers are used. This method reduces the computational overhead by performing heavy computations on the remote cloud servers [22]. But the state of cloud servers these days is highly centralized posing a single point of failure and security concerns. Infura was developed to enable decentralized applications to run without creating an Ethereum node. It extensively uses Amazon Web Services(AWS), so if some day amazon decides to shut down this service then the entire Infura will go down all at once. Cloud-based server offload is not a recommended solution as it is centralized in nature.
  - Side-Chain Offload: Sidechains run parallel to the main chain and are really effective in reducing the load o the main chain or we can say IoT devices. It enables the network to perform offload computations that do not require the whole blockchain to operate. Plasma, Liquid, Alpha, etc are some of the existing side chains [1].
  - Edge Computing: An enormous amount of latency is introduced into the system when deploying offload computations mainly due to the large distance between the cloud server and the operating node. Edge computing hence provides a mechanism to perform these computations closer to the end nodes or devices. Until 2017, Edge computing was termed as Mobile-Edge Computing which was later changed to Multiaccess Edge Computing (MEC) because of its wide range of benefits in the wired and wireless networks along with mobile networks. MEC allows offloading computation to either an edge computer or edge server since the close distance between the two decreases the communication latency significantly [22][23].

CONCLUSION

Blockchain is an emerging topic whose integration with IoT networks points toward the future aspects possible in various fields such as healthcare, supply chain, smart grids, and many more. IoT network faces many challenges such as reliable data transfer, device authentication, non-repudiation, and heterogeneity to which blockchain provides a solution to tackle all these problems. Blockchain being an immutable, decentralized ledger strengthens IoT networks. For instance, it is very critical to protect the patient's details as it is very sensitive and real-time transfer of data is very important as it is associated with a patient's life. But blockchain Itself suffers from some security issues such as double spending, etc. BIoT, Blockchain-based Internet of Things, despite being a powerful tool faces many challenges in its integration. An extensive and in-depth analysis of the challenges and their solutions in the integration of public blockchains, like Zilliqa, Elrond Blockchain, Monero, and IoT. Challenges related to the integration of Blockchain with IoT are storage, computational power, scalability, and efficiency of energy. Storage size, Scalability, computational resources, security, and energy efficiency, are the challenges that blockchains need to solve before successfully integrating blockchain with IoT. Various solutions have been presented in this paper to overcome these challenges. Resource-Constrained IoT Devices,that have very limited CPU utilities and small storage capacities so that these devices could perform at maximum efficiency at the expense of minimal energy/power. Due to their limited storage and low computational power, these IoT devices are incapable of storing substantial ledger sizes and high computational resources. The low-transaction rates of the existing blockchain pose additional hurdles inhibiting the blockchain from incorporating more and more IoT devices. So we also presented solutions even for computers with low computational resources like cloud offload, side-chain offload, and edge computing, etc. Finally, integration of IoT and Blockchain will strengthen the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:39:15 UTC from IEEE Xplore. Restrictions apply.

self-decision and trust issues among the communicating IoT devices extending the security it provides and it thus is making a transition from the traditional client-server architecture easier.

Future work will throw some light on the analysis of consensus algorithms and Merkle Tree acceptable for combination with resource-constrained IoT network to achieve better interoperability.

REFERENCES

[1] Khor, J. H., Sidorov, M., amp; Woon, P. Y. (2021). Public blockchains for resource-constrained IOT devices—a state-of-theart survey. IEEE Internet of Things Journal, 8(15), 11960–11982. https://doi.org/10.1109/jiot.2021.3069120

[2] Ferrag, M. A., Derdour, M., Mukherjee, M., Derhab, A., Maglaras, L., amp; Janicke, H. (2019). Blockchain technologies for the internet of things: Research issues and challenges. IEEE Internet of Things Journal, 6(2), 2188–2204. https://doi.org/10.1109/jiot.2018.2882794

[3] Dai, H.-N., Zheng, Z., amp; Zhang, Y. (2019). Blockchain for internet of things: A survey. IEEE Internet of Things Journal, 6(5), 8076–8094. https://doi.org/10.1109/jiot.2019.2920987

[4] A. Kumar, P. S. Mehra, G. Gupta, and M. Sharma, Enhanced block Playfair Cipher, vol. 115. 2013, doi:10.1007/978-3-642-37949-9_60.

[5] Devibala, A. (2019). A survey on security issues in IOT for Blockchain Healthcare. 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT). https://doi.org/10.1109/icecct.2019.8869253

[6] Xu, L. D., Lu, Y., amp; Li, L. (2021). Embedding blockchain technology into IOT for security: A survey. IEEE Internet of Things Journal, 8(13), 10452–10473. https://doi.org/10.1109/jiot.2021.3060508

[7] P. Gupta, P. Raj, S. Tiwari, P. Kumari, and P. S. Mehra, "Energy Efficient Diagonal Based Clustering Protocol in Wireless Sensor Network," SSRN Electronic Journal, Apr. 2020, doi: 10.2139/ssrn.3565781.

[8] Sasikala, N., Sundaram, B. M., Biswas, S., Sai Nikhil, A., amp; Rohith, V. S. (2022). Survey of latest technologies on decentralized applications using blockchain. 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS). https://doi.org/10.1109/icais53314.2022.9742768

[9] Meneghello, F., Calore, M., Zucchetto, D., Polese, M., amp; Zanella, A. (2019). IOT: Internet of threats? A survey of practical security vulnerabilities in real IOT devices. IEEE Internet of Things Journal, 6(5), 8182–8201. https://doi.org/10.1109/jiot.2019.2935189 [10] Conti, M., Sandeep Kumar, E., Lal, C., amp; Ruj, S. (2018). A survey on security and privacy issues of bitcoin. IEEE Communications Surveys amp; Tutorials, 20(4), 3416–3452. https://doi.org/10.1109/comst.2018.2842460

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

[12] Malik, S., Dedeoglu, V., Kanhere, S. S., amp; Jurdak, R. (2019). Trustchain: Trust Management in Blockchain and IOT supported supply chains. 2019 IEEE International Conference on Blockchain (Blockchain). https://doi.org/10.1109/blockchain.2019.00032

[13] Li, H., Lu, R., Zhou, L., Yang, B., amp; Shen, X. (2014). An efficient Merkle-tree-based Authentication Scheme for smart grid. IEEE Systems Journal, 8(2), 655–663. https://doi.org/10.1109/jsyst.2013.2271537

[14] A. Banafa. IoT and Blockchain: Challenges and Risks, Datafloq. Accessed: Apr. 3, 2020. [Online]. Available: https://datafloq.com/read/iotand-blockchain-challenges-and-risks/3797

[15] Zilliqa. (2017). The Zilliqa Techinical Whitepaper. [Online]. Available: https://docs.zilliqa.com/whitepaper.pdf

[16] Elrond. (2018). Elrond—A Highly Scalable Public Blockchain via Adaptive State Sharding and Secure Proof of Stake. [Online]. Available: https://elrond.com/files/Elrond Whitepaper EN.pdf

[17] A Scalable Vaue Transfer Protocol for the Digital Economy, Elrond, Los Angeles, CA, USA, 2018. [Online]. Available: https://elrond.com/files/Elrond DeckEN.pdf

[18] A. Back et al. (2014). Enabling Blockchain Innovations With Pegged Sidechains. [Online]. Available: https://blockstream.com/sidechains.pdf

[19] J.-P. Buntinx. Pro's and Con's on Bitcoin Block Pruning.Accessed: May 9, 2020. [Online]. Available: https://news.bitcoin.com/pros-and-cons-onbitcoin-block-pruning/

[20] J. Ehrenhofer, Monero Adds Blockchain Pruning and Improves Transaction Efficiency, Monero. Istanbul, Turkey, 2019. [Online]. Available: https://ww.getmonero.org/2019/02/01/pruning.html

[21] S. Popov, The Tangle, vol. 1, 3rd ed. Berlin, Germany: IOTA Foundation, 2017.

[22] A. K. Dwivedi, P. S. Mehra, O. Pal, M. N. Doja, and B. Alam, "EETSP: Energy-efficient two-stage routing protocol for wireless sensor network-assisted Internet of Things," International Journal of Communication Systems, vol. 34, no. 17, p. e4965, Nov. 2021, doi: 10.1002/DAC.4965

[23] F. Xu, F. Yang, C. Zhao, and C. Fang, "Edge computing and caching based blockchain IoT network," in Proc. 1st IEEE Int. Conf. Hot Inf. Centric Netw. (HotICN), Aug. 2018, pp. 238–239, doi: 10.1109/HOTICN.2018.8606001.

[24] H. Mayer. (2016). ECDSA Security in Bitcoin and Ethereum: A Research Survey. [Online]. Available: http://blog.coinfabrik.com/wpcontent/uploads/2016/06/ECDSASecurity-in-Bitcoin-and-Ethereum-aResearch-Survey.pdf

[25] .P. S. Mehra, "LBECR: load balanced, efficient clustering and routing protocol for sustainable internet of things in smart cities," Journal of Ambient Intelligence and Humanized Computing, Jan. 2022, doi: 10.1007/s12652-022-03706-2.

[26] S. O'Dea. (2019). Smart City IoT Active Connections in the EI 2016, 2019, 2022, and 2025. [Online]. Available: https://www.statista.com/statistics/691843/smart-city-iot-active-connections-in-the-eu/

[27] A. Palai, M. Vora, and A. Shah, "Empowering light nodes in blockchains with block summarization," in Proc. 9th IFIP Int. Conf. New Technol. Mobility Security (NTMS), Feb. 2018, pp. 1–5, doi: 10.1109/NTMS.2018.8328735.

[28] Algorand. Algorand Network. Accessed: Mar. 28, 2020. [Online]. Available: https://algorand.foundation/network

[29] S. Li and L. Xu, Securing the Internet of Things. Cambridge, MA, USA: Syngress Publ., 2017.

[30] S. Li, L. Xu, and S. Zhao, "The Internet of Things: A survey," Informat. Syst. Front., vol. 17, no. 2, pp. 243–259, 2015.

[31] A. Whitmore, A. Agarwal, and L. D. Xu, "The Internet of Things— A survey of topics and trends," Informat. Syst. Front., vol. 17, no. 2, pp. 261–274, 2015.

[32] Y. Lu, "Industry 4.0: A survey on technologies, applications and open research issues," J. Ind. Inf. Integr., vol. 6, pp. 1–10, Jun. 2017.

[33] Zhuang, P., Zamir, T., amp; Liang, H. (2021). Blockchain for Cybersecurity in smart grid: A comprehensive survey. IEEE Transactions on Industrial Informatics, 17(1), 3–19. https://doi.org/10.1109/tii.2020.2998479

[34] R. Rivest, "RFC 1321: The MD5 message-digest algorithm," in Internet Activities Board. Gaithersburg, MD, USA: National Institute of Standards and Technology, 1992.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:39:15 UTC from IEEE Xplore. Restrictions apply.

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-022-13837-5

TRACK 2: MEDICAL APPLICATIONS OF MULTIMEDIA

A survey on recent developments in diabetic retinopathy detection through integration of deep learning

Shalini Agarwal[1] · Aruna Bhat[1]

Received: 26 October 2021 / Revised: 31 May 2022 / Accepted: 6 September 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Diabetes, nowadays, is a very common disease throughout the world among people of all ages. The higher level of blood sugar in the blood more often leads to Damage to blood vessels in the retina leads to blindness if left untreated or undetected on time. Diabetic Retinopathy (DR) screening programs like retinal fundus image analysis help ophthalmologists deal with some visual impairment problems. Computer-Aided Diagnosis aims to detect the severity of DR as early as possible so that it can be handled before the occurrence of any irreversible vision loss. With the help of many advancements in artificial intelligence techniques, a highly efficient and accurate system can be designed to help medical professionals automatically diagnose DR at an early stage without any special clinical resources. This paper conducts a thorough investigation into several recent frameworks proposed based on machine learning and deep learning networks to classify non-proliferative diabetic retinopathy, exudates, hemorrhages, and micro aneurysms. Several promising pre-trained deep learning model to classify DR stages exploited by researchers, and also investigated Transfer learning on pre-trained GoogLeNet and AlexNet, VGG etc. models. It is observed that almost all public and private data sets widely available for research are imbalanced. To alleviate these issues, generative adversarial networks (GANs) and their variants were also used to generate label-preserving data. In this study, the authors also list the recently proposed GAN-based frameworks and their impact on model performance. The paper concludes with the current challenges and future directions in the early and accurate classification of the severity level of DR.

Keywords Diabetic retinopathy · Deep learning · GAN

✉ Aruna Bhat
aruna.bhat@dtu.ac.in

Shalini Agarwal
shaliniagarwal_phdco2k19@dtu.ac.in

[1] Department of Computer Science & Engineering, Delhi Technological University, 110042 Delhi, India

Published online: 27 September 2022

1 Introduction

Diabetic Retinopathy is a retinal disorder. The retina is a spherical structure made up of thin layers of tissues at the eye's inner backward wall. Its job is to analyze visual information using specialized cells like rods and cones, which are photosensors in the eye. The macula is a dark, round region located in the retina's central part. The center point of the macula, also known as the fovea, is important for giving clear vision. The retinal tissue, the same as any other tissue in the body, gets blood supply through the body's vasculature. The optic disc, which also includes the optic cup, is a part of the retina that holds the optic nerve as well as all of the major arteries and veins that in and out of the eye [47]. Diabetes, now a day, is the third deadliest disease after cancer and heart disorders among all age groups of people worldwide. As per a recent study by the International Diabetes Federation (IDF), Diabetes mellitus (DM) is a universal eye disease that is also considered the 4th leading cause of death among people of all age groups, and the frequency is supposed to increase by 7.7% by the end of the year 2030 [87, 95]. Diabetic Retinopathy is a retinal disease that is prevalent in patients suffering from diabetes. The blood vessel that nourishes the retina is blocked due to too much sugar level in the blood. It also blocks the light that passes through the optical nerves and the patients suffering from this disease can also have loss of vision if left undetected and untreated within time [85]. The new retina blood vessels did not develop properly and can also leaks easily again, which leads to DR. At the initial stage there is mild vision problem such as blurred vision, fluctuating vision, and bleary-eyed coloured vision but consequently, it may lead to loss of vision [104] The Diabetic Retinopathy eye can be seen in Fig. 1e.

Two forms of diabetes are linked to DR. Early DR symptoms are generally seen in those 75–95% of patients who suffers from type 1 diabetes for around 15 years. Patients with type 2 diabetes have around a 60% prevalence rate among those who have had the disease for more

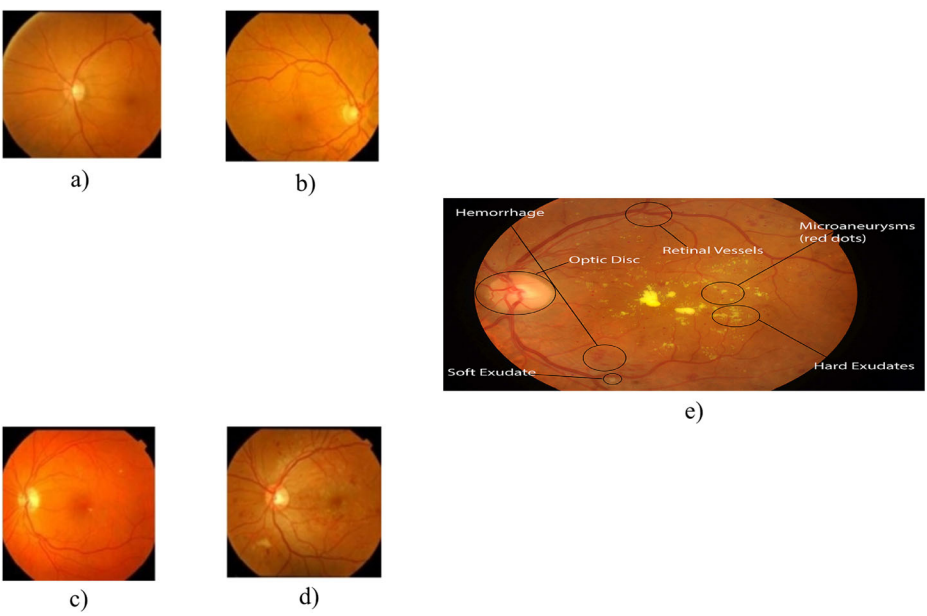


Fig. 1 **a** to **e** Healthy Retina, Mild NPDR, Moderate NPDR, Severe NPDR, PDR (includes all lesions that leads to severity like Hemorrhage, microaneurysms, soft and hard exudates.)

than 16 years. DR and its stages are diagnosed in around 80% of the people worldwide suffering from diabetes for at least ten years [47, 90]. The rapid increase in DR goes undetected due to the lack of evident alarming signs of DR before it leads to extreme vision loss. Early detection, proper medication, and routine care can help to reduce the severity of DR. DR can be diagnosed using a common visual diagnostic tool, digital fundus imaging. According to recent clinical research, precautionary visual tests should be conducted on diabetic patients at regular intervals, lifelong. Patients with none or mild DR should be advised to have an annual screening, while those with severe retinopathy should be advised to have a fundus examination every six months. With the rapid growth of patients diagnosed/suffering with DR, these bulk retinal tests have become a notable workload and difficult to manage job for ophthalmologist. It has been discovered that almost half of the population suffering from diabetes in the United States have been screened for DR eye disease on an annual basis. This number is alarmingly low in France with just 32% of diabetic patients being screened for DR. The rapid growth of patients suffering from DR, has posed a challenge to ophthalmologists with an increased requirement for retinal scans. In the early stages of DR, one may not experience any symptoms. As the disease progresses, there is an advent of symptoms like dark strings or spots floating in the vision(floater), blurred vision, fluctuating vision, and color vision impairment. In the later stages of the disease, there are noticeable dark areas in the vision meaning complete loss of vision [8, 11]. The presence of various forms of lesions on a retina image is used to detect DR. Microaneurysm (MA), hemorrhages (HM), soft exudates, and hard exudates (EX) or change in blood vessels [85] are the lesions shown in Fig. 1e. The earliest symptom of DR is microaneurysm (MA), which occurs as small red round spots occur at the retina because of weakness in the retina vessel's walls. There are sharp margins and the scale is less than 125 m. Hemorrhages (HM) are around 125 m diameter large spots on the retina and have an irregular margin. Flame (superficial HM) and blot (deep HM) are the two forms of HM (deeper HM). Plasma leakage triggers hard exudates, which appear as bright yellow spots on the retina. Soft exudates (also known as cotton wool) look like a white spot on the retina caused by nerve fiber swelling. They have very sharp margins and are generally located in the retina's outer layers. The shape of exudates is either Oval or round. MA and HM appear as red lesions, while soft and hard exudates appear as bright lesions (EX) [62]. Diabetic retinopathy is broadly classified into 5 types according to the feature present:

1. Pre-Proliferative Retinopathy: No DR means healthy retina.
2. Mild Non-Proliferative Retinopathy: Only microaneurysms.
3. Moderate Non-Proliferative Retinopathy (NPDR): Microaneurysms and at Least one out of Hemorrhage, Cotton wool spots and Hard Exudates.
4. Severe Proliferative Retinopathy: Any of the below

    a) Prominent intra-retinal microvascular abnormality.
    b) Definite venous beading in 2 quadrants.
    c) NO sign of PDR.
5. Proliferative Retinopathy (PDR): Neovascularization or vitreous/pre-retinal hemorrhage

Early DR This is the most common form of DR, also known as non-proliferative DR (NPDR). At this point, the walls of the retina's blood vessels become weaker, and new blood vessels do not form. When the microaneurysms bulge out from the walls of blood vessels in the retina,

blood spills out. The central part of the retina, known as the macula, will swell in this stage; this condition must be handled as soon as possible.

Severe DR This stage of extreme DR is often referred to as proliferative DR (PDR). When damaged blood vessels are blocked, new blood vessels expand in an unstructured manner. The retina begins to detach from the back of your eye as a result of these new scar tissues stimulation. Glaucoma is a condition in which the nerve that transmits signals from the eye is weakened. It is a resulting condition from long ignored severe DR. As a consequence, vision loss occurs [27, 39].

NPDR is a form of diabetic retinopathy that involves blood vessel leakage into the retina. Consequently, the retina becomes inflamed and moist. At the NPDR level, retinal anomalies such as microaneurysms, exudates (hard and soft), and hemorrhages can be detected. Exudate are the basic symptoms of DR, it presents in two forms: Hard and Soft. hard exudates are looked like yellow waxy patches in the retina while Soft exudates can be seen as a white or light yellow area with distracted edges [39]. Hemorrhages are marked by the texture of blood spots on the retina. One of the most severe causes of DR is the presence of exudates in retinal fundus images. MAs are also the first symptoms of a widening of the retinal vasculature [1]. Figure 1 depicts a typical retina and various levels of DR. Figure 1a shows the healthy retina image. Figure 1b depicts the mild PDR in this only microaneurysms is visible. Figure 1c shows the symptoms of Moderate NPDR where Microaneurysms and at Least one out of Hemorrhage, Cotton wool spots and Hard Exudates are visible. Figure 1d shows Severe NPDR in which prominent intra-retinal microvascular abnormality exist and finally Fig. 1e shows the PDR image in which all lesions leads to blindness exists.

Manual diagnosis is more expensive and time-consuming than automated methods for DR detection [73]. Manual diagnosis has a higher risk of misdiagnosis and takes longer than automated procedures. This study examines contemporary DR automated methods that move away from classical machine learning and toward deep learning, as well as the usage of GAN to generate more accurate data for training and classification of DR. This paper is organized as follows: Section 2 list the literature resources like all the available public and private retina fundus image data-sets and trends. Section 3 briefly discusses research methodologies and work done so far in recent years and also presents various performance measures. Based on that study Section 4 enlist some research gaps for future work. At last Section 5 concluded the study.

2 Literature resources

2.1 Available datasets

Fundus images have been utilized to diagnose a variety of retinal illnesses in recent years, including cataracts, glaucoma, keratoconus, Diabetic Retinopathy, and age-related macular degeneration. Images of the retinal fundus are taken with a fundus camera. This camera captures images of the inner retina, macula, blood vessels, posterior pole, and optic disc. In medical diagnostics, image capturing is a crucial stage. There are so many authentic public and private datasets exploited by researchers like HEI-MED [69], MESIDOR [55], MESIDOR 2 [3], FIRE [38], STARE [81], E-OPHTHA-MA and E-OPHTHA-EX [26], IDRID [41], ROC

(Retinopathy Online Challenge), DR1 DR2 [56], DIARETB0 and DIARETB1 (Image-Ret) [22], KAGGLE [44], DRIVE [25], ARIA [10], DriDB [24], Moor Field eye hospital data [50], Singapore National Diabetic Retinopathy Program Private data [45], HUMP [1], CLEOPATRA [84] and many more private data sets are available. The details of these data sets are listed in Table 1. This table has Datasets Name, whether its public or private data set, Number of fundus eye images, resolution of these images, and the device type used to capture these images. Some data sets of images were captured at different camera settings at different angles.

To the best, KAGGLE [44], is used by so many researchers these days as it is the largest publicly available dataset for DR research. It is publicly provided for research by EyePACS which is an open data source for all eye-related disease detection. In the year 2015 EyePACS held a competition for automated detection and classification of all 5 DR severity classes such as Proliferative DR, Severe DR, Moderate DR, Mild DR, and No DR. On analysing the data ,it is found that the data is imbalanced, means the number of images for each classes are not equally distributed. This count imbalance creates problems while evaluating the performance of various machine learning models. To deal with this problem [101] used a subset of labelled images from this dataset for classification using Deep CNN to maintain the balance among different classes. They utilized only 1000 images out of approximately 88,000 fundus images present in the dataset. Out of 1000, 800 were used for training of the model, and all the remaining 200 were used for testing and performance evaluation. Some studies have utilized just exclusive datasets, while others have utilized a mix of private and publicly available datasets. an exclusive dataset called CLEOPATRA was used by Tan et al. [84] in their proposal to detect and classify various eye abnormalities using deep learning, whereas Wang et al. [89] employed numerous datasets that include two publicly available datasets (ROC and DIARETDB1) and one private dataset from Moor-fields Eye Hospital London. For all of the public datasets that were imbalanced, if used to classify the DR severity levels, the authors had to use different augmentation strategies or select nearly equal amounts of a sub-set of data for training and testing. This could result in the loss of critical information, and because of these methods of augmenting data, there's a probability that they may be inaccurate. Hence, a balanced dataset would result in better accuracy in results. Figure 2 shows some samples of images from different datasets. We can see the diversity in images in different dataset. This problem of diversity should be addressed.

2.2 Research trends

The need of detecting Diabetic Retinopathy at early stages have emerged in recent years. At the non-proliferative stage, the disease is defined by the existence of microaneurysms before progressing to the proliferative stage, which is marked by neovascularization. Retinal experts work to discover this disease early so that it can be treated within time before it causes significant and irreversible vision loss. This result is designing an automated system using eye fundus images. Figure 3 shows the trend according to PubMed in the interest taken by so many researchers worldwide to work with eye fundus images to detect the various types of lesions caused by diabetes. According to studies, almost all type I diabetes patients and 60% of those with type II diabetes will be affected by proliferative DR within twenty years after diagnosis. There is a tremendous growth in research in DR after 2015. With the advancement in artificial intelligence and machine learning algorithms, Fig. 4 shows the use of deep learning mechanisms to widely increase the accuracy of detecting and classifying the severity levels of DR [82].

Table 1 List of Public and Private datasets

Dataset Name	Availability	Number of images	Resolutions	Camera
HEI-MED	Public	115 abnormal, 54 healthy	2196×1958	45 fold-view with Zeiss visucam-Pro fundus
MESSIDOR	Public	1200	1440×960, 2240×1488 2304×1536	3CCD
MESSIDOR 2	Public	1784	Different resolutions	Non-mydriatic Topcon
FIRE	Public	134	------	Nidek AFC-210
STARE	Public	400	605×700	TOP-CON-TRV 50 with 35 fold-view.
E-OPHTHA- EX E-OPHTHA- MA	Public	E-OPHTHA- EX: 47 exudated & 35 no lesion E-OPHTHA- MA: 233 No lesion & 148 small hemorrhage and microaneurysms.	2048×1360	OPHDIAT
ROC (Retinopathy Online Challenge)	Public	1200	768×576 1058×1061 1389×1383	Canon
DR1 & DR2	Public	DR1- 234 DR2- 520	---	---
DIARETB0 & DIARETB1 (Image-Ret)	Public	DIARETB0 -130 DIARETB1- 89	1500×1152	50-fold view
KAGGLE	Public	88,702 (Training- 35,126, Testing- 53,576)	Different resolutions	----
IDRID	Public	Segmentation: 81 Disease Grading: :516 Localization: 516	----	----
DRIVE	Public	Training: 20, Testing: 20	786×584	3CCD
ARIA	Public	142	---	Zeiss FF450
DriDB	Public	50	---	University of Zagreb
Moor Field eye hospital data	Private	21,536	---	-----
Singapore National DR Program private dataset	Private	197,085	----	----
Hump	Private	250	---	Hospital of Puerta Del Mar, Spain
CLEOPATRA	Private	298	----	Collected by 15 hospitals in the UK

2.3 Metrics for model performance evaluation

There are many pre-processing techniques to be used to enhance eye fundus images before giving them to any machine learning model for feature selection and classification. There are various measures to be used to check the strength of these models. Evaluation in medical imaging is generally based on two scenarios: the correct location of the lesions present in the image and the evaluated level of correctness and confidentiality. There are 4 terms generally

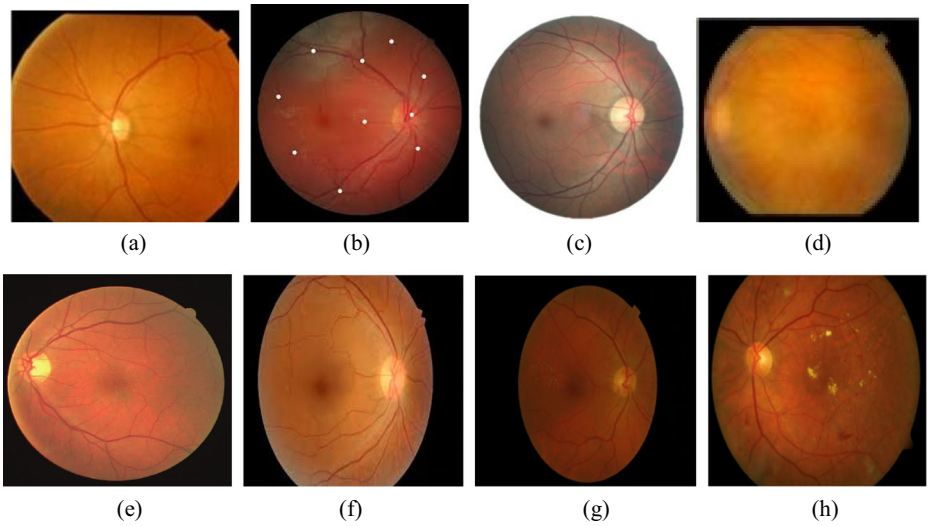


Fig. 2 **a** to **h** Sample retina fundus images: MESSIDOR, FIRE, KAGGLE, STARE, DRIVE, IRDID, E-OPHTHA, DIARETDB1 illustrating diversity among datasets

used in the evaluation matric: True Negative(TN), True positive(TP), False Positive(FP), and False Negative(FN). TP denotes the number of correctly classified positive lesions. FP is the number of negative lesions classified as positive. TN refers to the number of correctly classified negative lesions and FN is the number of positive lesions incorrectly classified as negative. Specificity and sensitivity are to be measured based on the values of these four terms. PSNR denotes peak signal to noise ratio (The log decibels values), high PSNR shows the quality of the process images is better than the real images [9]. Specificity, sensitivity is the

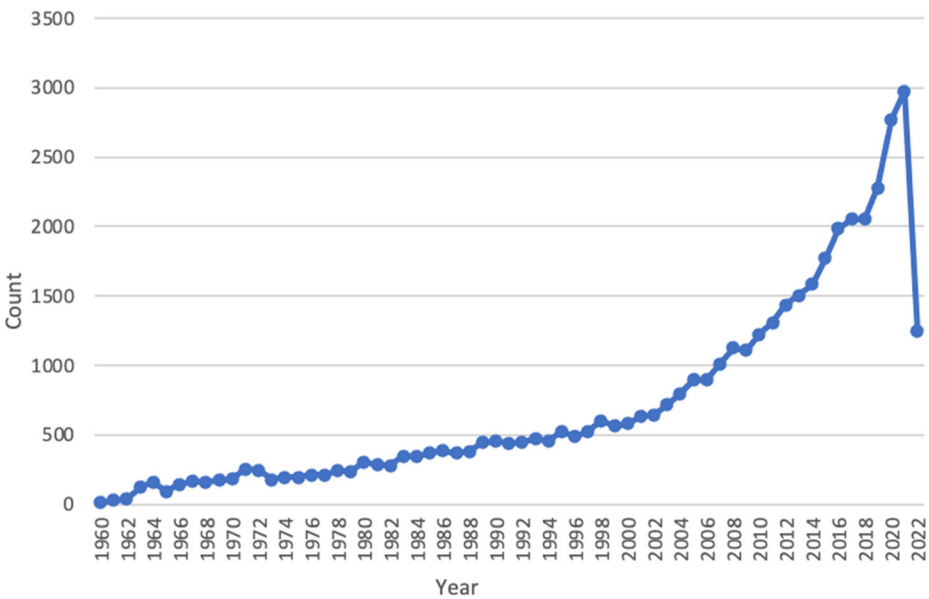


Fig. 3 Keyword 'Diabetic Retinopathy' publication Trend according to PubMed

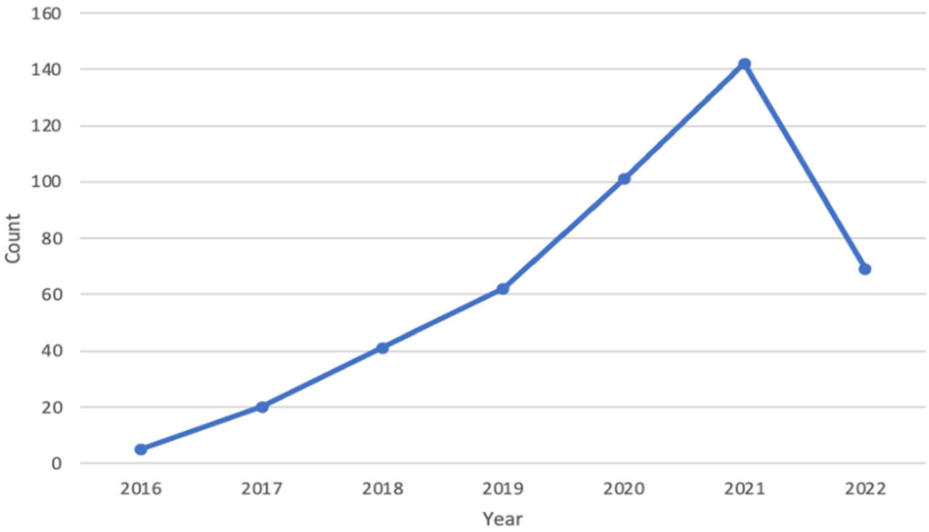


Fig. 4 Keyword 'Diabetic Retinopathy' & 'Deep Learning' publication Trend by year according to PubMed

commonly used measure of the efficiency of classification models designed for medical images. Listed here are some common formulas for evaluation to help the researchers to compare various model performances.

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP} \times 100 \tag{1}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{2}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{Precision} = \frac{TP}{FP + TP} \tag{4}$$

$$\text{PSNR} = 10log_{10}\frac{(PeakValue)^2}{MSE} \tag{5}$$

$$\text{F} - \text{Score} = \frac{2*\text{Recall}*\text{Precision}}{\text{Recall} + \text{Precision}} \tag{6}$$

$$\text{Kappa score} = \frac{ACC - ACC_{Prob}}{1 - ACC_{Prob}} \tag{7}$$

3 Methodology

Various automated DR detection models have been developed till now by researchers to diagnose MA, HM, and soft and hard exudates (EX) at early stages. Still, there is a scope to improve the accuracy of these existing CAD systems because so many small retinal lesions are also left undetected. Figure 5 shows the common model architecture to read and classify the eye fundus image into various DR severity stages as Pre-Proliferative Retinopathy, Mild NPR, Moderate NPDR, Severe Proliferative Retinopathy and PDR as shown in Fig. 1.

3.1 Input eye image corpus

There are various public and private retina color image datasets available for research generated by numerous scientific communities. Some of the private clinical institutions also provide data for research to facilitate research on designing automated diabetic retinopathy classification tools. We have covered details of these datasets in Section 2.1. These available image corpus may be noisy and needed to be pre-processed and refined for further feature extraction.

3.2 Image preprocessing & augmentation

Generally, there is no uniform standard of image appearance and quality among all datasets of DR as shown in Fig. 3. every particular image has different illumination, the dynamic broad range of aspect ratio, extra borders, different camera light, focus and color settings. These non-uniform input samples reduce the performance of classification as it is tough to detect very small lesions like EX, MA, HM, and hard and soft cotton wool in different illumination. In this step, some image pre-processing techniques are to be applied for extracting more distinctive features from the image and clean the irrelevant noises from the image. For better visualization, there are so many pre-processing methods exploited by researchers till now. There are two ways to work with the input image: one with the different color channel extraction(RGB) and another is by converting it into gray level first. Generally, the green channel is extracted as it contains more information than the other two (Red and Blue). In [106] green channel is extracted for MAs detection. Contrast limited adaptive histogram equalization is then applied in that extracted green channel for contrast enhancement and to improve the visibility of hidden features.

Gaussian filtering is the well-known filter to remove noise from the image to make it smooth. Some researchers also remove the blood vessels and mask optical disc to improve results as they can be misclassified as lesions. In [68] a color RGB fundus image is converted

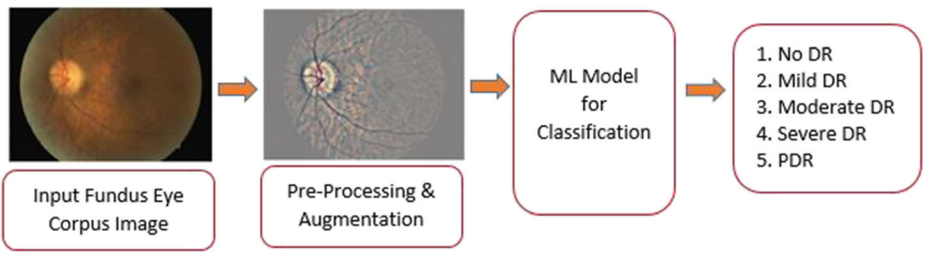


Fig. 5 Typical model architecture for Diabetic Retinopathy detection

to a lightless channel J (contains only grey information) of CIECAM02 color space model, then to improve varying contrast and remove odd illumination, non-linear contrast enhancement is applied to this J channel. Image resizing is another pre-processing method in which image resolution is compromised to converse the computation power. It is generally used with machine learning or deep learning models. This task is carried out in almost 25% of the models. In [23] first green channel is extracted, contrast enhancement using histogram equalization, then every image is resized to 512 × 512 pixels to make the input corpus into a standardized form. Image augmentation was also applied. Image augmentation is the AI process to make a balance between different classes of images. It is used when there is a huge imbalance in the dataset. In this process, images are flipped, rotated, translated, or mirrored to produce more instances of images of that class that have fewer numbers in comparison to others. In [67] image augmentation is applied as: rotation in angle 0 to 360, scaling 85–115%, translation from − 10px to + 10px, horizontal flipping and contrast modification ranges from 60 to 167%.

3.3 DR detection & segmentation techniques

There are various retinal characteristics and features that can be used to classify the level of DR. Various techniques have been used till now to detect and extract these features. These techniques are still the focus of the comprehensive research.

3.3.1 Blood vessel segmentation

In fundus images, extraction of retinal blood vessels is critical because it provides information that helps ophthalmologists diagnose the disorders such as age-related macular degeneration, glaucoma, diabetic retinopathy etc. Blood vessel segmentation can also be used to diagnose brain and heart diseases caused by aberrant changes in the retina's vascular system. The non-invasiveness of fundus imaging makes this type of segmentation more significant. Figure 6a shows blood vessel extraction and segmentation.

Blood Vessels segmented help to reduce the false positives rate while detecting dark lesions on the retina and also track changes in microvascular while detecting PDR. false positives occur because of the similar pixel intensities between Blood Vessels and certain lesion types. PDR is defined by angiogenesis or the formation of new retinal BVs that lack structural integrity. Generally, these vessels are unable to function normally, and they are more likely to burst or cause irreversible harm. The various image processing techniques used to segment blood vessels are summarized in Table 2.

Kedir et al. (2018) [5] proposed a robust approach for detection of longitudinal retinal changes due to small red lesions. This methods normalized images by reducing illumination variations that improves the contrast of small retinal features. a simple and effective blobness measure is also proposed to detect spatio-temporal retinal changes. support vector machine classifier used to identify several intensity and shape features. Experimentally this method is able to detect retinal changes due to small red lesions with a sensitivity of 80% at an average false positive rate of 1 and 2.5 lesions per eye on small and large fields-of-view of the retina, respectively.

A new hierarchical image matting model proposed by Zhun Fan et al. (2019) [28] and achieved segmentation accuracy of 95.7%, 96.0% and 95.1% on STARE, DRIVE and CHASE_DB1 datasets respectively. This matting model generates specific trimap

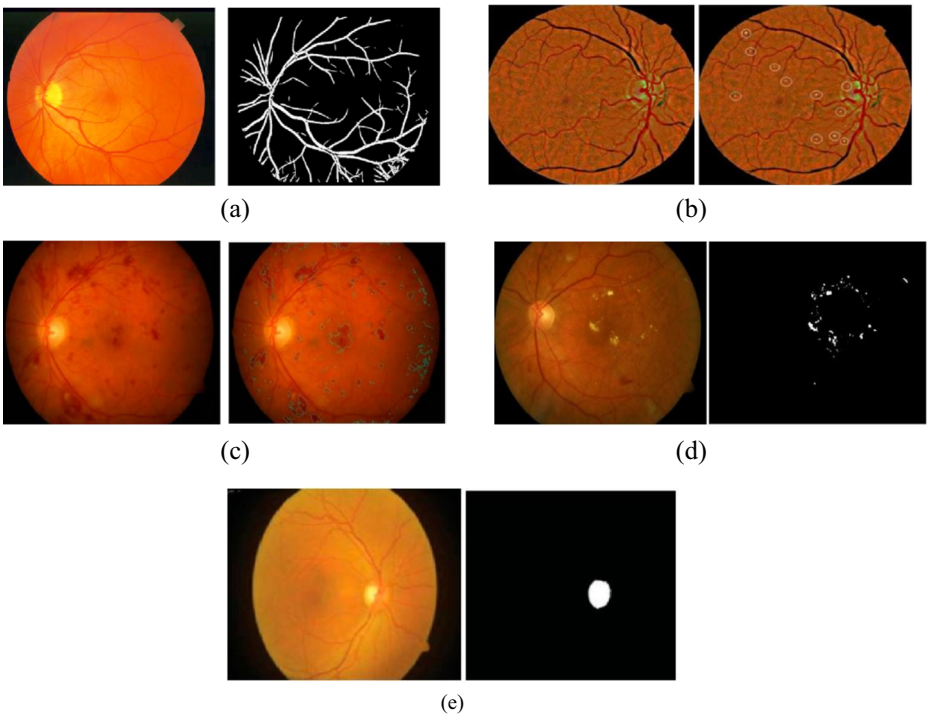


Fig. 6 **a** to **e** DR Segmentation Techniques: Blood vessel segmentation, MA Detection, Hemorrhages detection, Exudates Detection, Optic disc and optic cup segmentation

automatically which separates the image into 3 regions: background, foreground and unknown. This technique have low calculation time(average 15.74 s, 10.72 s 50.71 s respectively on STARE, DRIVE and CHASE_DB1 datasets) and also outperform many traditional learning models.

Sonali et al. (2019) [70] proposed a noise removal and contrast enhancement approach. This approach works on image enhancement and de noising issues by using used integration of various filters and CLAHE technique. This method achieved 1.19% improvement in SSIM, 7.85% improvement in PSNR, 0.12% improvement in CoC and 1.28% improvement in EPI comparative to other state of the art method till date.

Henry et al. (2019) [50] focuses on information loss due to quality of images. Different key performance indicators and methods applied on common datasets and new PixelBNN model introduced for automating the segmentation of fundus image morphologies. This model tested on STARE, DRIVE and CHASE_DB1 and achieved 90%, 91,% and 89% accuracy respectively.

A new cascade classification framework introduced by Xiaohong et al. (2019) [93] for blood vessel segmentation. This framework can handle typically challenging retinal blood vessel structures. This model envelops a set of computationally efficient Mahalanobis distance classifiers to form a highly nonlinear decision. Also this framework trainind by a one pass feed forward process. 95.41–96.40% vessel segmentation accuracy achieved on STARE, DRIVE and CHASE_DB1 datasets.

Table 2 Details of different types of segmentation and detection techniques

Ref.	Year	Category	Database	Technique Used	Performance
[5]	2018	Blood vessel segmentation	Rotterdam Eye Hospital	Multi-scale blob detection Approach	Sensitivity: 98%
[28]	2019	Blood vessel segmentation	STARE, DRIVE, CHASE_DB1	Hierarchical image matting model	Accuracy: 95.7%, 96%, 95.1%
[70]	2019	Blood vessel segmentation	STARE	CLAHE	EPI: 1.28% PSNR:7.85%, SSIM: 1.19%, CoC: 0.12%
[50]	2019	Blood vessel segmentation	STARE, DRIVE, CHASE_DB1	PixelBNN deep method	Accuracy: 90%, 91%, 89%
[93]	2019	Blood vessel segmentation	STARE, DRIVE, CHASE_DB1	Cascade classification framework	Accuracy:95.41%, 96.40% and 96.3%
[80]	2019	Blood vessel segmentation	DRIVE, STARE, CHASE_DB1, and HRF	strided-CNN, PCA, Morphological operation, class balancing loss function	Sensitivity: 87%, 80.8%, 88.6%, 82.9% Accuracy: 95.6%, 95.4%, 97.6%, 96.2%
[43]	2019	Blood vessel segmentation	DRIVE, STARE, CHASE_DB1, and HRF	Deformable U-Net framework (DUNet)	Accuracy: 95.66%, 96.41%, 96.10%, 96.51%
[60]	2019	Blood vessel segmentation	CHASE_DB1	scale-space approximated (SSA) CNN (SSANet)	AUC: 99.16
[29]	2020	Blood vessel segmentation	DRIVE, STARE	Cross-connected CNN (CCNet)	Sensitivity: 76.25%, 77.09% Accuracy: 95.28%,96.33%
[86]	2021	Blood vessel segmentation	DRIVE, STARE, CHASE_DB1	classical edge detectors(Robert, Prewitt, Canny, Sobel, and morphological filters), Neural Network	Accuracy: 94.80%, 95.48%, 94.52%, sensitivity: 0.7352, 0.7265, 0.7279, specificity : 0.9775, 0.9759,0.9658
[40]	2022	Blood vessel segmentation	DRIVE, CHASE, STARE	Dilated U-net	Accuracy:96.80%, 97.01% and 96.94%
[36]	2018	microaneurysms detection	DIARETDB1	Two-step DCNN; GoogLeNet, double-ring filter, Gabor filter, Hessian matrix;	sensitivity: 84%
[14]	2018	microaneurysms detection	DIARETDB1	PCA, Radial Basis function, SVM	Accuracy: 98.5, F-Measure: 92.6
[74]	2019	microaneurysms detection	Messidor, DiaretDB1	Ant colony algorithm, BV Removed	Dice index values of 0.81–0.90
[20]	2019	microaneurysms detection	DIARETDB1	Cascaded Multi-sieving CNN (MSCNN), hybrid text image features	Recall: 87.8, Precision: 99.7, Accuracy: 96.1, F-Score: 93.4
[21]	2019	microaneurysms detection	1140 retinal color photographs	Histogram, Local Binary Pattern(LBP), discrete orthonormal	Sensitivity: 96.5 Specificity:95.3

Table 2 (continued)

Ref.	Year	Category	Database	Technique Used	Performance
				Stockwell transform, and statistical features	
[65]	2020	microaneurysms detection	IDRID-2019	mutual information (MI), Semantic segmentation, Deep CNN, Maximum matching filter response (MFR), and maximum Gaussian answer laplacian (LoG) in the 2-D function space	Sensitivity: 97.4% for dark lesions & 96.8% for bright lesions
[100]	2021	microaneurysms detection	E-Ophtha-MA	MSRNet, MS-EfficientNet	Sensitivity: 0.715
[53]	2021	microaneurysms detection	Retinopathy Online Challenge dataset, E-ophtha-MA	deep convolutional encoder-decoder network, smooth dice loss	Sensitivity: 0.7819
[63]	2021	microaneurysms detection	E-ophtha, DiaRetDB1	image normalization, gradient weighting, shade correction, hessian-based filter,- Otsu thresholding, Grey wolf optimizer, GLCM	Accuracy: 99.10, 99.90
[105]	2021	microaneurysms detection	IDRiD	Unet++network, dicecross loss (DCEL)	AUC: 0.6144
[16]	2022	microaneurysms detection	Real time images, E-ophtha,	Monte Carlo statistical analysis	Analysis
[94]	2022	microaneurysms detection	FFA image cohort	FC-DenseNet, FFA images, Histogram Stretching, Gaussian Filtering	Accuracy: 94.19% Precision: 88.40% Recall: 89.70% F1-score : 88.98%
[75]	2018	Hemorrhages Detection	EyePACs, MESSIDOR	multiple CNN-based detectors, hybrid lesion-based algo-rithms	AUC: 95%, 98%
[7]	2019	Hemorrhages Detection	GopRetina, ImageRet	**segmentation of vessels**: Gabor Transformation, **removal of the vessels**: Top-hat transformation, firefly algorithm (FFA) and particle swarm opti-mization algorithm (PSOA)-based itera-tive thresholding, sup-port vector machine (SVM) and linear re-gression (LR)-based classifier	specificity: 91.4%, sensitivity: 96.7%, accuracy: 94.1%
[17]	2019	Hemorrhages Detection	DiaretDB0, DiaretDB1	Random Forest classifier, shape and region-based features	Overall Accuracy: 93.58%, Sensitivity: 73.13, Specificity: 95.26
[99]	2019		DIARETDB		specificity: 82%,

Table 2 (continued)

Ref.	Year	Category	Database	Technique Used	Performance
		Hemorrhages Detection		2D Gaussian fitting, CLAHE, background estimation and watershed segmentation	sensitivity: 100%, Accuracy: 95.42%
[64]	2020	Hemorrhages Detection	FIRE	CLAHE, FCM, NN	Accuracy: 99%
[37]	2021	Hemorrhages Detection	private dataset: DR-D, which contains 420 images	incremental segmentation approach, knowledge distillation, **Probability map alignment**	With the integration experiments, we get the optimal values of the model-based weight to be 0.7
[66]	2021	Hemorrhages Detection	IDRiD	FCN-8 architecture, modified CLAHE	Specificity: 98.81%, sensitivity: 55.45%, accuracy: 98.25
[46]	2019	Exudates Detection	DIARETDB1, E-Ophtha	SVM, Discriminative Restricted Boltzmann Machines, and pre-trained network (ResNet-50)	Accuracy: 97.6%, Sensitivity: 99%
[97]	2019	Exudates Detection	DIARETDB1	mathematical morphology algorithm (MMA), morphology mean shift algorithm	Sensitivity: 98.40% Specificity: 98.13% Accuracy: 98.35%,
[54]	2019	Exudates Detection	DIARETDB1, DIARETDB0, HRF	Different morphological operations, CLAHE	Sensitivity:94.59%, Specificity: 88.46%
[52]	2019	Exudates, Hemorrhages and microaneurysms	Kaggle, DDR dataset	HED, DeepLab-v3+, Caffe framework and fine-tuned the weights pre-trained on the Pascal VOC dataset	Intersection over union (IoU) of MA: 0.0311/0.325, HE: 0.5634/0.3118, SE: 0.4359/0.2295
[57]	2020	Exudates, Hemorrhages, and microaneurysms	EyePACS dataset from Kaggle 2015 competition	HED, CNN, VGG-16, and VGG-19	sensitivity: 80%, accuracy: 82%, specificity: 82%, and AUC: 0.904
[76]	2020	Exudates, Hemorrhages, and microaneurysms	Messidor	Histogram based segmentation, Synergic deep learning (SDL) model	Accuracy: 99.28%, Sensitivity: 98.54%, Specificity: 99.38%
[92]	2019	OD segmentation	Collected CFI dataset of 2978 images	U-net model based on CNN to segment OD, vessel density map	Accuracy: 96.9%, Sensitivity: 93.9%
[12]	2018	OD segmentation	RIM-ONE	localization and segmentation of OD based on Circular Hough Transform (CHT) and localized active contour model (ACM), morphological closing and opening operations	Accuracy: 98%,
[30]	2019	OD segmentation	ORIGA	DL architecture: M- Net (U-shape convolutional	CDR: 89%

Table 2 (continued)

Ref.	Year	Category	Database	Technique Used	Performance
				network, multi-scale input layer, multi-label loss function, and side-output layer), Polar Transformation to improve segmentation	
[61]	2019	OD segmentation	OCT dataset	fully convolutional networks (FCN) with Gaussian Processes, GP based Regression	AUC: 98%
[35]	2021	OD segmentation	IDRiD, RIMONE, DRISHTI-GS, DRIVE	DRNet, Residual skip connection	mIoU: 0.845 on IDRiD, 0.901 on RIMONE, 0.933 on DRISHTI-GS, 0.920 on DRIVE

A deep strided CNN for accurate segmentation of BV (especially the tine one) is proposed by Toufique et al. (2019) [80]. PCA based along with morphological mapping steps used for pre-processing. features from the encoder part and the decoder part are concatenated using Skip connection to make tiny vessel's edge sharper. Class balancing loss function used to optimize image quality. This model outperforms other models with a sensitivity of 0.87, 0.808, 0.886 and 0.829, and accuracies of 95.6%, 95.4%, 97.6% and 96.2% on DRIVE, STARE, CHASE_DB1 and HRF respectively.

Qiangguo et al. (2019) [43] proposed a Deformable U-Net, which works on the retinal vessels' local features. DUNet combines low level and high level features for precise localization and extract context information. This model also scales the retinal vessel shapes according to the shape scale of standard one. DRIVE, STARE, CHASE_DB1 and HRF were used and achieved global accuracy of 95.66%, 96.41%, 96.10% and 96.51% respectively. Kyoung jin et al. (2019) [60] proposed a simple multi-scale CNN structure for bblood vessel segmentation. Scale-Space theory incorporate to propose this simple multi scale structure using upsampling. scale-space approximated CNN (SSANet) also incorporate residual blocks. This Net achieves state-of-the-art AUC value of 0.9916 for the ROC curve.

S. Feng et al.(2020) [29], proposed a cross connected CNN for retinal tree segmentation automatically. This Net fuse multilevel features with the help of cross connection between primary and secondary path. Also trained, tested and validated with full green channel images directly. Experiments performed on DRIVE and STARE datasets and achieved accuracy of 95.28% and 96.33% respectively. Also they claimed that this algorithm has faster (0.063 s) computing speed comparative to other GPU based algorithms.

Beaudelaire et al. (2021) [86] used classical edge detectors and the neural network for retina blood vessel segmentation. laplacian of Gaussian, Robert, Prewitt, Canny, Sobel, and morphological filters are used for producing features vector. This feature vector serves as an input to train a cascade feed-forward neural network. a high quality vessel probability map shows while experimenting on DRIVE, STARE and CHASE datasets. This architecture claimed to be robust with diverse luminosity conditions and varying widths of vessel. 94.80%, 95.48% and 94.52% accuracy, 0.7352, 0.7265, and 0.7279 sensitivity, 0.9775, 0.9759, and 0.9658 specificity achieved on DRIVE, STARE and CHASE_DB1 datasets respectively.

Hussain et al. (2022) [40] proposed a dilated U-Net architecture for better transfer, accurate classification and are very much effective at capturing dynamically complex retina blood vessels. This convolution is main constitutions of encoder, decoder block and skip connections. The input image is scale down and concatenated with convolutional block outputs in decoder path to ensure feature transfer. Up-sampled output obtained from multioutput module from each decoder block. This improved U-Net architecture achieved accuracy of 96.80%, 96.94%, and 97.01% and sensitivity of 88.37%, 82.63%, and 87.13% on DRIVE, STARE, and CHASE datasets respectively. These metrics shows that this architecture perform much higher than original U-Net and also robust to noise.

3.3.2 Microaneurysms detection

Tiny aneurysms are fundamental and micro indicators of diabetic retinopathy in fundus imaging. Micro aneurysms are caused by the leakage of retinal blood vessels, which eventually turn into circular red spots on the retina. these are very small size marks. Due to the low contrast between the retinal surface and the red lesion, microaneurysms are generally most difficult to see when it comes to DR. The graphical representation of microaneurysm identification is shown in Fig. 6b. There are various classifiers used to identify the red lesions and classify the abnormalities present in the retina. Xia et al. [100] present a novel approach using a combination of random forest classifier and K mean to develop CAD for accurate classification. Accuracy with this approach was 93.58% which is considerably improved than Naive Bayes which reaches only to 83.63% accuracy. Various deep learning approaches were also developed by researchers. Table 2 shows the methods used in recent years along with the accuracy achieved.

Hatanaka et al. (2018) [36] Combines 3 existing types of microaneurysms detector as shape index based on the Hessian matrix, the double-ring filter and gabor filter. They also used 2 step Deep convolution neural networks for training and testing. The first DCNN step is for MA detection and second is for false positives (FPs) reduction. DIARETDB1 database used to test and validate and 84% sensitivity achieved.

Wen cao et al. (2018) [14] used small 25 × 25 pixel patches from segmented from DIARETDB1 dataset. Raw pixel intensities from these patches used as an input to random forest, SVM and neural network. principal component analysis used for dimensionality reduction. According to author, in comparison with traditional machine learning methods, this proposed deep learning-based MA detection method outperformed with AUC improved from 0.962 to 0.985 and F-measure improved from 0.913 to 0.926.

V. Deepa et al. (2019) [21] proposes a method based on discrete orthonormal Stockwell transform and statistical features for differentiating between the effected and normal retina. Total of 24 feature extracted from these two different approaches. 8 best known classifiers compared with this proposed method for detection of MAs on 1140 color retinal images. This proposed method shows promising results with sensitivity of 96.5% specificity 95.3%.

Linf Dai et al. (2019) [20] presents multi-sieving convolutional neural network (clinical report guided), which drag a small amount of supervised information from these clinical reports to identify the microaneurysm regions. This process used image-to-text mapping in the feature space. In highly unbalanced classification problem, these MA regions interleaved with retina image information for multi sieving deep mining. This Cascaded Multi-sieving CNN validated on DIARETDB1 dataset and achieved 99.7% precision and 87.8% recall.

T. Selcuk et al. (2019) [74] used segmentation using ant colony algorithm in place of conventional image processing techniques. Retinal vascular structure extracted from retinal images before segmentation. random walker, watershed, Maximum entropy, K-Means and region growing, these 5 different image processing and clustering algorithms also implemented for performance comparison. Dice and Jaccard similarity index values used for similarity matching between automatic and manually MA segmentation done by clinical experts. Dice index values of 0.81–0.90 achieved.

Lifeng Qiao et al. (2020) [65] proposed a semantic segmentation algorithm for classification of normal and affected retina images. CNN with deep learning as a core component along with GPU perform detection and segmentation with low latency inference and achieve high performance. Maximum matching filter response (MFR), and maximum Gaussian answer laplacian (LoG) in the 2-D function space used on IDRID-2019 dataset. This architecture achieved Sensitivity 97.4% for dark lesions & 96.8% for bright lesions.

Haiying et al. (2021) [100] proposed a 2 stage model, Segmentation and classification. a multi-scale residual network (MSRNet) designed to extract all the necessary information from the MAs with large size variance. In second stage, a multi-scale efficient network(MS-EfficientNet) developed which can discriminate the false MAs from the MA patience detected by MSRNet. E-Ophtha dataset is used and this model outperform with an sensitivity average of 0.715 with highly robust and effective segmentation.

Yinhan et al. (2021) [53] proposed a novel end to end deep convolutional encoder- decoder network for detection of MA in color retina fundus images. The smooth dice loss, which is a weighted dice loss is used to put more focus on misclassified Mas. Finally, to produce an accurate probability map for MA detection, an activation function with a long tail is used. Retinopathy Online Challenge dataset and the e-ophtha-MA dataset used for experiments and this simple and effective method guarantees the shortening of test time. Sensitivity of 0.7819 achieved.

Manohar et al. (2021) [63] introduce a six phase model. In first phase lower light enhancement, shaded correction, and gradient weighting is used to improve the visibility of E-ophtha, and DiaRetDB1 datasets. In next stage, the hessian-based filter, and Otsu thresholding used to eliminate blood vessel regions from the MA regions. Grey wolf optimizer is used for predicting accuracy of the these segmented MA regions. After that For feature extraction, Gray Level Co-occurrence Matrix (GLCM) is used. Modified K Nearest Neighbour (MKNN) is used for classification of microaneurysms and non-microaneurysms regions. 99.10% and 99.90% of accuracy achieved on e-ophtha, and DiaRetDB1 respectively.

Guihao et al. (2021) [105] use Unet + + network as backbone for small lesion segmentation. As we know that microaneurysm datasets are limited so a mixed pre-processing is used to reconstruct the whole dataset and to handle dataset imbalance problem, dicecross loss (DCEL) used. 0.6144 AUC achieved which is better than the first ranker of IDRiD challenge.

Choong et al. (2022) [16] works on the accuracy and consistency of MA detection models. Otsu's multi-thresholding has been used and perform very well in many applications but this paper analyse the Monte Carlo statistical analysis using synthetic images. These retinal images varied such as: stages of DR, image parameters like MA size, intensity differences between different MAs and blood vessels (BVs), as well as measurement noise. To guarantee accurate and consistent DR detection, Real retinal images tested. The final results shows that if the intensity difference between BV and MA is 50% and above and MAs size is ranging from 5 to 20 pixel then the results will be accurate and consistent.

Zhenhua et al. (2022) [94] improve the detection of MA in fluorescein fundus angiography (FFA) images using FC-DenseNet. Histogram Stretching and Gaussian Filtering used for FFA image enhancement. MAs-FC-DenseNet compared with other similar FC-DenseNet models: FC-DenseNet56 and FC-DenseNet67. The final result suggested that MAs-FC-DenseNet has higher values of accuracy, MPA, precision, recall, F1 score and mean intersection 99.97%, 94.19%, 88.40%, 89.70%, 88.98%, and 90.14%, respectively over union over other models. It is also a reliable model for fast and accurate detection of MAs. All these experiments performed on FFA image cohort which was constructed in the collaboration of Nanjing Medical University, Affiliated Hospital of Soochow University, and Huai'An First People's Hospital. This dataset contains 1200 FFA images (768 × 868 pixels) from 1200 eyes of DR patients from the age range 31 to 81, who underwent FFA in three hospitals from Aug, 2015 to Dec, 2019.

3.3.3 Hemorrhages detection

The presence of hemorrhages in fundus pictures is an early-stage symptom and considered as a serious cause of DR. In this case, precise and very early detection of hemorrhage is become essential. This can help to treat patients on time before any irreversible effects. Hemorrhages are found in both the superficial and deep retinas. The shallow retina has linear-shaped hemorrhages and bright red in color, while the deep retina has encircled-shaped hemorrhages and dark red in color. Various approaches are used to identify hemorrhages, which are detailed in this section. The detection of hemorrhages is depicted in Fig. 6c. To change the brightness generally HSV color shape with a non-linear curve is used. Gamma correction is used to highlight brown areas that represent Hemorrhages. There are numerous methods designed for hemorrhage detection, listed in Table 2.

Abhay et al. (2018) [75] investigates the relative susceptibility of image based and hybrid lesion based algorithms. This paper focus on slightly different adversarial images generated from referable diabetic retinopathy (rDR) and lead to misclassification. There are two type of algorithms: one image based where input is retinal image and generate a blackbox kind of classification output, another is hybrid lesion-based algorithms that used multiple CNN-based detectors. This paper compare diagnostic accuracy of these above two types of algorithm on rDR generated adversarial images.

Kemal et al. (2019) [7] proposed an firefly and particle swarm optimization based iterative thresholding approach that can be used to detect of hemorrhages. In this framework: image enhancement done by basic image pre-processing methods, Gabor and Top-hat transformation used for vessel segmentation and discriminate number of hemorrhages regions. Further pixel count in these segmented regions done by firefly algorithm and PSOA based iterative thresholding. SVM and LR based classifier used to detect hemorrhages. 91.4% specificity, 96.7% sensitivity and 94.1% accuracy achieved.

Amrita et al. (2019) [17] used Random forest classifier for acute diagnosis of HM. After pre-processing, K means based algorithm used to group bright and dark lesion clusters. DiaretDB0 and DiaretDB1 datasets used in which ground truth is also listed. Shape and region based 14 features extracted and this classifier compared with neive bayes classifier for accuracy. This proposed framework shows accuracy of 93.58%.

In Jun Wu et al. (2019) [99], 2D Gaussian fitting based method proposed. Brightness correction and CLAHE used for pre-processing. watershed segmentation and background estimation used to detect hemorrhages and after that 2D Gaussian fitting and human visual

characteristics used to extract visual features of HM. 219 images of DIARETDB database is used to evaluate this framework and shows 100% sensitivity, 82% specificity and 95.42% accuracy at image level and 94.01% sensitivity, 90.30% positive predictive value at lesion level.

HA Purwanithami et al. (2020) [64] proposed a method for Hemorrhage Detection based on CLAHE and FCM. FIRE dataset is used and first images are resized to 150 × 200 and then green channel is choosen because blood vessel contrast in green channel is better than others. CLAHE is used for image enhancement and FCM is a segmentation technique that cluster data based on distance between input and nearest cluster. Neural Network is further used for classification based on generated features. An experiment conducted on 100 testing image with 2 classes: Hemorrhage, Non-Hemorrhage. As an outcome only one image incorrectly classified which conclud that accuracy reached 99%.

Wanji He et al. (2021) [37] works on the problem of pixel wise annotations, which is labour intensive when we required to add new data and classes frequently. A novel incremental segmentation approach is proposed in this paper to distill the knowledge of the previous trained model to improve the classification power of current model. Remarkably, we propose various approaches working on the class-based alignment of the probability maps of the current and the previous model, accounting for the difference between the background classes of the two models. They designed their own private dataset: DR-D, which contains 420 images (consist of optic disc, arc spot, HE and exudate) with high-quality pixel-level annotations on the classes.

Qomariah et al. (2021) [66] propose a modification of the FCN-8 architecture by adding a modified convolution and batch normalization identity mapping to enrich the features of the for hemorrhage segmentation process. Extended version of CLAHE used to enhance image. This proposed method achieved 98.81% specificity, 55.45% sensitivity, and 98.25 accuracy with IDRiD dataset.

3.3.4 Exudates detection

Exudates are a very common early symptom of Diabetic Retinopathy. The accumulation of more protein and lipid in the retina is known as exudate. The retina is commonly detected by very bright, reflex, white, or milky lesions. It indicates a greater risk of retinal edema and higher vascular permeability. It is not considered a potential visual threat but an indication of excess fluid accumulation in the retina. However, they appear towards the macula's center and are regarded as a hazard to the lesion. These are divided into two categories: soft and hard exudates. Soft ones also known as cotton wool spots are depicted as light yellow or white areas with distracting borders, while hard exudates are depicted as yellow waxy patches on the retina. General Exudate detection is shown in Fig. 6d. Exudates detected early may aid patients in receiving correct and timely therapy for diabetic retinopathy. The method used by researchers is listed in Table 2 in brief.

Dulanji et al. (2019) [54] proposed a method to automatically detect exudates from retinal images. DIARETDB1 dataset used which contains 89 images with array size of 1500 × 1152 pixels. First they standardized these images to scale to 640 × 480 pixels to improve processing efficiency. DIARETDB0 and HRF dataset also used for further evolution of this algorithm. Green channel extracted from these normalized images because of high in contrast. Further optical disk location has been identified more accurately and it is important factor for the removal of false positives. No ML techniques used in this framework.

Quang et al. (2020) [57] exploits various machine learning models such as CNN, VGG-16 and VGG-19 for automatic classification of DR. input image analysed for varied illumination and field of view. This method achieves sensitivity 80%, accuracy 82%, specificity 82%, and AUC 0.904 for 4 stage DR classification.

K. Shankar et al. (2020) [76] focus on DR classification on the basis of severity level using a deep learning model. In pre-processing, tradition image enhancement method used to remove noise that exist in the edges. In Segmentation step, Histogram based segmentation used to extract important regions that contains lesions. Further, Synergic Deep Learning (SDL) model applied for classification of stages of DR on Messidor dataset and 99.28 accuracy, 98.54 sensitivity and 99.38 specificity achieved.

Parham et al. (2019) [46] also investigates deep learning techniques and compare both supervised and unsupervised models for classification accuracy. Also to improve the performance pre-trained residual networks (ResNet-50) and Discriminative Restricted Boltzmann machine experimented on 2 publicly available datasets: DIARETDB1 and E-Ophtha. ResNet-50 with SVM outperformed other networks with an accuracy of 98% and sensitivity of 0.99.

Kittipol et al. (2019) [97] develop a new and improved framework to detect exudates lesions in color images by using a morphology mean shift algorithm. First they normalize the retina images, enhance the contrast, remove edge noises and then localize OD. Classification using mathematical morphology algorithm (MMA) procedure is applied to keep only exudates pixels. Based on comparison of images and corresponding ground truth available in DIARETDB1 and a local dataset for exudates detection, this proposed methos achieved average sensitivity 98.40%, specificity 98.13%, and accuracy 98.35%.

3.3.5 Optic disc-optic cup segmentation

Clinically, the optic disc is an important fundus element in fundus imaging, and it is an early indicator of the macula, optic cup, fovea, and retinal vessel detection. OD is a nearly circular shape area behind the eye where the optic nerve is surrounded by retinal nerve fibres. It is located in front of the optic nerves when the eyes enter the brain this is why it is sometimes referred to as the optic nerve head (ONH). When light enters the eye and reaches the center of the OD, it does not respond and cannot convert it into true sensory impulses. It also does not send any messages to the brain for explication. The center and a disc contour are defined by the optic disc boundary for localization and segmentation. For the most part, geometric and morphological approaches were utilized to extract the optic disc in the literature, listed briefly in Table 2. The segmentation of OD and OC is shown in Fig. 6e.

Shreenidhi et al. (2018) [12] suggests a novel framework for OD segmentation using image processing methods, localized active contour model and Circular Hough Transform. Morphological opening and closing operations applied for noise removal and removal of gray scale gaps. RIM-ONE database is used and shows a remarkable detection of OD region with 98% accuracy.

Huazhu et al. (2019) [30] proposed a deep learning architecture: M-Net for OD and OC segmentation in multi label system. This M-Net have multi scale input layer, a U-Net CNN and one side output layer. Multi-level loss function used to generate final segmentation map. Polar transformation introduced further to improve segmentation accuracy. ORIGA dataset used.

Lei Wang et al. (2019) [92] proposed a coarse to fine DL framework known as an update on U-Net. This model trained separately on color retina images and their grey scale vessel density

maps. The results combined using an overlap strategy to identify Disc candidate region, which will feed back to U-Net for further segmentation. This method experimented over 2978 test images from local and 6 publicly available datasets, outperform with intersection over union of 89.1% and dice similarity coefficient of 93.9% over U-Net.

Hasan et al. (2021) [35] proposed an encoder- decoder network: DRNet for OD and Fovea segmentation. for compensating the spatial information lost due to pooling in the encoder, they proposed a residual skip connection. RSC is different from old skip connections as they directly concatenate low level feature maps from beginning layers of encoder. DRISHTI-GS, IDRiD, RIMONE, and DRIVE datasets used for testing OD segmentation. Mean Intersection over Union (mIoU) of 0.845 on IDRiD, 0.901 on RIMONE, 0.933 on DRISHTI-GS, and 0.920 on DRIVE achieved.

Deep learning, which uses neural networks to calculate thousands of mathematical equations with millions of parameters, is commonly utilized for image categorization applications. Recent DR detection research has generally focused on the development of new algorithms for retinal images that are mostly impacted by occlusion, blurriness, refraction, and lighting fluctuations. Some latest DL-based techniques irrespective of individual segmentation techniques for DR detection will be discussed further. Figure 5 also represents the number of research papers published applying deep learning for DR since 2016 and its increasing with the advancements in technology day by day [4, 71, 88]

In 2015, The EyePACS retina image dataset was launched publically with 35,126 training and 53,576 testing images in the Kaggle competition [33], which was one of the major breakthroughs in DR stage classification. It drew the attention of computer vision researchers and data scientists to collaboratively work from all over the world, and many deep learning algorithms for detecting DR were presented. Gulshan at al. [34] develop a CNN-based deep learning model called DCNN for DR image classification which was trained on EyePACS data set having 9963 images as well as the MESSIDOR data set having 1748 images. they trained an Inception-v3 architecture [83] with a total of 128,175 images from both data sets and achieved significant results in 2 operating cut schedules with high sensitivity and specificity. Li et al. [51] also uses Inception-v3 network on 19,233 color retinal fundus images and optimizes using 10-fold cross-validation. R. Gargeya et al. [31] proposed a customized CNN that classifies into two classes as healthy retina or unhealthy (any DR stage). this network trained over 75,137 images and for external validation, tested over two public datasets MESSIDOR-2 and E-Ophtha. They used AUC to measure precision-recall trade-off and achieved a 0.94 AUC score on MESSIDOR 2 and 0.95 on E-Ophtha. C. Lam et al. [48] used a pre-trained deep learning model to classify DR stages and investigated Transfer learning on pre-trained GoogLeNet and AlexNet models from ImageNet. They used image augmentation techniques to prevent overfitting. Overall 74.5% accuracy was achieved on testing over the ImageNet dataset. Y. Wu et al. [98] experiment transfer learning on image data taken from the Kaggle data set. the preprocessed data are given to pre-trained VGG19, ResNet, and InceptionV3 networks. This method achieves significantly better results. He et al. [37] used NN architecture which is based on MobileNets. 16,798 images of the left and right eye were used. Tensor-flow is used in MobileNets training with the help of asynchronous gradient descent, batch norm and ReLU used to prove non-linearity after each layer. 73.3% accuracy with 74.5% sensitivity was achieved while testing on 1000 random samples. J. H. Tan at al. [84] used 10-layer CNN for Automatic segmentation of all small lesions like micro-aneurysms, exudates, and hemorrhages on the CLEOPATRA database. odd-numbered images used for training and even for testing. because of highly skewed data, training is performed in two

stages, and standard backpropagation and stochastic gradient are used for training in both stages. To prevent overfitting, regularization is used. this model achieved 62.57%, 95.72%, 46.06%, and 87.58% sensitivity for hemorrhages, background, micro-aneurysms, and exudates respectively. K. Adem [6] works to detect exudates using CNN and Hough transform applied in fundus images for the first time. adaptive histogram equalization and Canny edge detection algorithm were applied for image enhancement and segment OD regions from images. A success rate of 99.18%, 98.53% and 99.17% was reported for exudates detection from DrimDB, DiaretDB1 and DiaretDB0 public datasets respectively

Chen et al. [15] developed a structured model pruning library to modify models in place without adding mask layer. They show that CNN is often hindered by high storage and computational requirements. SMP is proved to be promising to alleviate these requirements and also can improve memory usage and speed on TPU while preserving accuracy for small datasets like CIFAR10. They have experimented TensorFlow 2 implementation of VGG16 pre-trained on ImageNet and CIFAR-10 datasets to study accuracy-efficiency trade-off. X. Wang et al. [91] used CNN with 166 high-quality images selected from the Kaggle data set to avoid the problem of overfitting due to an imbalanced data set. Three different CNN models: AlexNet, VGG16, and InceptionNet V3, were trained. InceptionNet V3 achieves significantly higher accuracy as 63.23%

3.4 Generative Adversarial Network in DR

Diabetic retinopathy can be detected and classified using fundus eye images. fundus images are a very effective and non-invasive method to detect severity levels of DR. A fundus camera is used to capture images. very large and balanced (almost equal number of images of all classes) dataset, as well as, robust deep learning algorithms are required for accurate lesion diagnosis using deep learning models. The continuous minimization of validation errors with the training error is a key aspect in constructing a robust deep learning model. Designing Deep learning models, that are both resilient and highly accurate But are not cost and resource effective The discrepancy between the performance of models on training labelled data, and the results on real data; that has never been seen before is referred to as generalizability in deep learning models. This problem can be improved by using a broad and balanced dataset that covers a wide range of problem descriptions. Over-fitting occurs when a model is trained using a bad dataset. However, accessing a significant and suitable amount of data in some fields of research is not simple. Patient confidentiality, skilled professionals to annotate the data, and the effort to collect the data are all factors to consider when working with medical datasets [77]. We require a large number of labelled retina images with a balanced distribution of different levels of disease, appropriate resolution, and no aberrant pixels in the case of diabetic retinopathy. Even the largest publicly available dataset for diabetic retinopathy detection Kaggle [44] is an imbalance as approximately 73% of total data is belongs to level 0. Image-based Data augmentation is one of the solutions to address this issue, in which some geometric transformations like image resizing, cropping, rotating, flipping, scaling, etc. are applied [77] to expand and enhance the data. These basic image transformations are not able to produce clean and valid data that can contribute to increasing the performance of deep learning models [102].

The deep learning-based generative model shows promising results in contrast to image-based augmentation. Goodfellow et al. [32] designed a special neural network The Generative Adversarial Network (GAN), having two generative and discriminative models can be trained

simultaneously as shown in Fig. 7. The generative model competes with an opponent (discriminator) by attempting to generate a comparable sample from input data, while the discriminator attempts to detect whether the generated sample came from the data distribution or the model distribution. GAN is generally used to synthesize imbalanced data and has gained much popularity, especially in medical imaging since its development by [32]. Welander et al. [96] used GAN in generating dataset as an unpaired image to image translation like generating MR images from CT images or generating multi-modality images set by using a single modality. Nie et al. [58] used GAN for medical image synthesis.

In medical imaging, GAN can be applied in two ways. The first is to concentrate on the generative feature, which aids in the discovery of hidden patterns in input photos and the generation of new images based on that input. The second is to concentrate on GAN's discriminative feature, which may be used to learn the pattern of normal images and detect abnormalities when fresh images are presented [102]. There are various GAN models to be applied in Diabetic retinopathy detection including Boundary Equilibrium GAN (BEGAN), Deep Convolutional GAN (DCGAN), Diabetic Retinopathy GAN (DRGAN), Conditional GAN (CGAN), Symmetric Equilibrium (SEGAN), Spare GAN (SGAN), Semi-Supervised Learning GAN (SSL-GAN), CycleGAN. Further, these GAN models are divided into Generative focused and Discriminator focused. Generative-focused GAN mainly works on synthesis, feature segmentation, de-noising, and transfer while discriminator-focused GAN focus on classification and detection. The basic network model proposed by [32] for the synthesis of artificial images also known as vanilla GAN showed in Fig. 7. G is the generative model and D is the discriminative model. Generative model G generates samples directly reading inputs z from noise distribution p(z). The generator output is G(z) or Xg. Xg and the instance of real distribution Xr (Generated from real data $P_{data}(x)$) given as an input to discriminator model D. D produce a single value (the probability of generated sample) that decides whether this generated sample is real or fake. The generator gets rewarded as positive gradient learning from the discriminator when the generated image is very close to real and punishes when the image is different. D works as a binary classifier to differentiate between real and fake samples and generator G is a fully connected neural network to generate realistic image samples to confuse D. Mathematical Eq. 1 represents the models G and D [32].

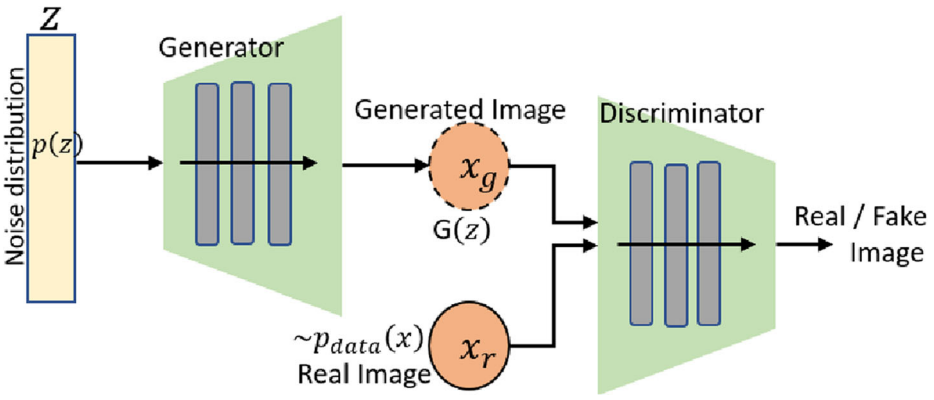


Fig. 7 Basic GAN Architecture showing Generator and Discriminator

$$Min_G Max_D V(D,G) = E_{x_r \sim P_{Data(x)}}[\log D(x_r)] + E_{x_g \sim P_z(z)}[1 - \log D(G(z))] \quad (8)$$

Appan et al. [78] produced a model based on GAN to generate synthetic images of hemorrhages with severity level 4. They presented a GAN model with two parallel generator networks, one with masked vessel trees and the other with masked lesions. Each network's output (which is based on U-net) is coupled with another U-net-based architecture to build a retinal image with all the necessary lesions. To differentiate the real image and the one generated by the generator, they used discriminator networks with five-layer convolutional layers. It stated that the synthetic images with Hemorrhage inclusion in the training set resulted in the improvement in sensitivity from 20 to 25% with just expert marked data. Costa et al. [18] proposed a model in which a vessel segmentation technique was used to pair true eye images with respective binary vessel trees. CGAN is used for an image to image translation to generate more realistic images. Experiment results represent that the generated and original images are visually different even though they generates the same blood vessel tree. also, the images possess true image set qualities. Costa et al. [19] presents an architecture that used generated masks of vessel trees to synthesise color fundus images. To recreate a realistic retinal vascular tree, model trained an adversarial autoencoder network. The reconstructed vessel trees of the retina are translated to a color fundus image using a generator and discriminator pair. This approach can provide a realistic retinal fundus visual texture as well as an associated vascular tree. Burlina et al. [13] used progressively grown GAN ProGAN for synthesizing fundus images for age-related macular symptoms. ProGAN is another version of GAN to specially work on the limitation related to the low resolution of generated output images. It starts with an input of 2*2 pix for a simple network for training. with each next iteration, network layers will increase and the resolution will be doubled. With this approach, they have achieved a final 512*512 resolution. Nie et al. [59] proposed a pathological description that employs CNN's activated neurons in the DR detector to automatically translate the lesion onto the output images. Given the descriptor and retinal vascular tree, this allows for the creation of synthesized, fully controlled diseased retinal pictures using a GAN-based technique. The U-net network structure is used to create the generator and discriminator. They can modify the amount and location of pathological evidence on the generated photos. Zhou et al. [108] provided an end-to-end strategy for improving vessel segmentation performance that included multi-scale features refine blocks (MSFRB), symmetric equilibrium generative adversarial network (SEGAN), and an attention mechanism (AM). SEGAN was utilized to create symmetric pictures of the fundus with more realistic local detail. Instead of using the standard up-sampling procedure of generator, which generally compromises high resolution for the extraction of discriminative features, MSFRB was presented to maintain image resolution as well as preserve semantic features together. Inside MSFRB, AM is used to focus on elements that are advantageous to the task of creating maps. Yu et al. [103] use a new preprocessing pipeline Multiple-channels-multiple-landmarks (MCML), which aims to synthesise color retinal fundus images with the help of mixing vessel tree, optic disc, and optic cup pictures. On two public fundus image datasets (DRIVE and DRISHTI-GS), they compared single vessel mask input and MCML mask input using various Pix2pix and Cycle-GAN designs. In addition, a novel Pix2pix structure with a ResU-net generator has been built and compared to the previous models. In this study, a Pix2pix network with a ResU-net generator employing MCML and high-resolution paired images can generate good and realistic retinal images. It shows that this proposed MCML method has a lot of potential in glaucoma computer-aided

diagnosis based on retinal fundus images. Son et al. [79] suggested RetinaGANs, a method for precise retinal vascular and optic disc segmentation that is similar to CGAN. They provide a funduscopic image to a generator, which turns it into a probabilistic map of the vasculature and optic disc. The discriminator develops a probabilistic map from a funduscopic image to identify whether the probabilistic map is from a real source or the generator. To feed the discriminator, the generator will attempt to increase its performance. They found that retinal vascular segmentation improved significantly, but not in the optic disc.

The discriminator network in GANs is usually divided into two categories of focus: classification and detection. The procedure entails training a GAN model with only photos of patients without diabetes, such that the discriminator can detect irregularities when presented with photos with diabetes symptoms [102]. In [107] Sparsity-constrained Generative Adversarial Network (Sparse-GAN) was presented as a framework for detecting abnormalities utilizing only healthy fundus images in the training dataset. For the detection of abnormalities, they used image-to-image GAN [42] to reduce the effect of image noise such as speckle noise. They used an encoder to convert the image's construction into latent space and suggested a new Sparsity Regularization Net based on spare LSTM to regularize the sparsity of latent space. The proposed method's accuracy is comparable to that of supervised learning approaches. Among image-to-image GAN and AnoGAN, they get the highest AUC level in both tests and the validated set [72].

For the classification of anomalies in retinal pictures, Lecouat et al. [49] developed a patch-based semi-supervised deep learning utilizing GAN (SSL-GAN). Before starting the training, they separated the high-resolution retinal picture into equal patches. After that, each patch is predicted separately to create an image-level prediction from all of the patches. The SSL-GAN will use discriminatory behavior to detect the image's specific class, as well as whether the created image is fraudulent or real. With only 10–20 annotated training photos, the SSL-GAN attained an AUC of 64%, compared to 47% for supervised approaches.

4 Future scope

We have Achieved significant accuracy for DR lesion classification but still, there are some notable limitations to work on to further improve the performance. This section will list some of these research gaps.

1. The current deep learning models for DR severity diagnosis did not focus much on the location of different DR lesions. Studying the depth and location of DR lesions is also important for accurate clinical diagnosis. CNN architectures could improve the classification accuracy by paying attention to small MA as well as study of hemorrhages shapes.
2. Latest Deep learning models have achieved significantly high accuracy in classifying two classes as DR and non DR but still did not achieve good accuracy in detecting the type of affected lesion as well as all five DR stages. Also in DL, various pre-train transfer learning architectures like VGGNet, AlexNet, etc. were used recently but almost all are trained for objects like food, animal, generic objects, etc. These architectures are less suitable for achieving higher accuracy on real-time medical data classification.
3. Sometimes only image data is not sufficient for accurate prediction but investigating patient's metadata like duration of diabetes, type of diabetes, age, genetic information and other clinical data regarding diabetes also play an important role. fusion of all this information can result in a more accurate prediction.

4. As listed in Table 1, almost all dataset has less number of images which is imbalance in nature as we do not have an equal number of images for all stages of DR. For achieving higher accuracy using Deep learning models usually require very large balanced training dataset. Various GAN architectures proposed recently for the generation of images as discussed in Section 3.4 but still, no architecture seems to fit perfectly for all stage classification.
5. The retinal images in all datasets are different in light incident angle, dimensions, illumination, contrast, etc. A robust cross data verifiable model needed to be developed. Neural Style transfer models can be explored to efficiently handle domain shift problems.
6. All datasets captured clinical retina images high cost devised. A portable low cost smartphone based Diagnosis model has not been explored in which smart phone captured images can be used for early screening of DR. That can potentially reduce the cost and also promote awareness in remote areas.

5 Conclusion

This study presents the recent developments in automatic diabetic retinopathy detection at an early stage to prevent blindness. A computer-based screening system reduces the time and efforts taken to diagnose the severity level of DR by ophthalmologists. this study reviewed approximately 70 research papers. This study begins with the types of diabetic retinopathy and their severity levels according to the present features like new retinal blood vessels, hemorrhage(HA), exudates(EX), and microaneurysm(MA) in retina images. there are so many authentic public and private datasets available for researchers are listed in Section 2. This paper briefly explains the CAD system for diabetic retinopathy detection and the performance measures used. there are various traditional machine learning algorithms used to design CAD systems initially. Deep Learning models show significant improvements in perfect classification. in this paper, we list some of the recent advanced deep learning architectures used along with their performance measures. Various deep learning pretrained models like AlexNet, VGG16, and InceptionNet V3, GoogLeNet etc. applied in recent years for multiclass classification and also has achieved significant results. CNN and Hough transform applied in fundus images for the first time and also various researchers woks on basic edge detector algorithm along with DL models to generate a noise free blood vessel and tree. integrating two deep learning models: EyeNet and DenseNet, produce more accurately diagnosing of the healthy and other DR stage on 4 different benchmark datasets with average accuracy of 91.35 [2]. When it comes to dealing with the one important challenge of imbalanced data while working with a deep learning model, many medical imaging researchers have taken notice of GAN. Due to a large number of datasets with unbalanced data variety, has gained increased attention for the identification of diabetic retinopathy during the years 2017–2019. According to studies, GAN is capable of generating new images with DR symptoms to train a deep learning model for illness categorization. Lastly, this paper list some recently published GAN architectures used in Diabetic Retinopathy detection. Still Results from this whole studies show that average accuracy is approximately 91% for correct classification of all DR stages.

Funding This study is not funded by any agency.

Data availability The manuscript does not have any associated data which was generated.

Declarations

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

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

References

1. Abbas Q, Fondon I, Sarmiento A, Jiménez S, Alemany P (2017) Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features. Med Biol Eng Comput 55(11):1959–1974
2. AbdelMaksoud E, Barakat S, Elmogy M (2022) A computer-aided diagnosis system for detecting various diabetic retinopathy grades based on a hybrid deep learning technique. Med Biol Eng Comput:1–24
3. Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, Niemeijer M (2016) Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Investig Ophthalmol Vis Sci 57(13):5200–5206
4. Achanta SDM, Karthikeyan T, Kanna RV (2021) Wearable sensor based acoustic gait analysis using phase transition-based optimization algorithm on IoT. Int J Speech Technol. https://doi.org/10.1007/s10772-021-09893-1
5. Adal KM, Van Etten PG, Martinez JP, Rouwen KW, Vermeer KA, van Vliet LJ (2017) An automated system for the detection and classification of retinal changes due to red lesions in longitudinal fundus images. IEEE Trans Biomed Eng 65(6):1382–1390
6. Adem K (2018) Exudate detection for diabetic retinopathy with circular hough transformation and convolutional neural networks. Expert Syst Appl 114:289–295
7. Adem K, Hekim M, Demir S (2019) Detection of hemorrhage in retinal images using linear classifiers and iterative thresholding approaches based on firefly and particle swarm optimization algorithms. Turkish J Electr Eng Comput Sci 27(1):499–515
8. Alban M, Gilligan T (2016) Automated detection of diabetic retinopathy using fluorescein angiography photographs. Report of standford education
9. Amin J, Sharif M, Yasmin M (2016) A review on recent developments for detection of diabetic retinopathy. Scientifica 2016:6838976
10. Aria dataset. https://www.eyecharity.weebly.com/aria_online.html
11. Arunkumar R, Karthigaikumar P (2017) Multi-retinal disease classification by reduced deep earning features. Neural Comput Appl 28(2):329–334
12. Bhat SH, Kumar P (2019) Segmentation of optic disc by localized active contour model in retinal fundus image. In: Smart innovations in communication and computational sciences. Springer, pp 35–44
13. Burlina PM, Joshi N, Pacheco KD, Liu TA, Bressler NM (2019) Assessment of deep generative models for high-resolution synthetic retinal image generation of age-related macular degeneration. JAMA Ophthalmol 137(3):258–264
14. Cao W, Czarnek N, Shan J, Li L (2018) Microaneurysm detection using principal component analysis and machine learning methods. IEEE Trans Nanobiosci 17(3):191–198
15. Chen K, Franko K, Sang R (2021) Structured model pruning of convolutional networks on tensor processing units. arXiv preprint arXiv:2107.04191
16. Choong K, Heng et al (2022) Performance analysis of multi-level thresholding for microaneurysm detection. Multimed Tools Appl:1–20
17. Chowdhury A, Roy T, Chatterjee, Banerjee S (2019) A random forest classifier-based approach in the detection of abnormalities in the retina. Med Biol Eng Comput 57(1):193–203
18. Costa P, Galdran A, Meyer MI, Mendonça AM, Campilho A (2017) Adversarial synthesis of retinal images from vessel trees. In: International conference image analysis and recognition. Springer, pp 516–523
19. Costa P, Galdran A, Meyer MI, Niemeijer M, Abràmoff M, Mendonça AM (2017) Campilho end-to-end adversarial retinal image synthesis. IEEE Trans Med Imaging 37(3):781–791
20. Dai L, Fang R, Li H, Hou X, Sheng B, Wu Q, Jia W (2018) Clinical report guided retinal microaneurysm detection with multi-sieving deep learning. IEEE Trans Med Imaging 37(5):1149–1161
21. Deepa V, Kumar CS, Andrews SS (2019) Automated detection of microaneurysms using stockwell transform and statistical features. IET Image Proc 13(8):1341–1348
22. Diaretdb0 dataset. http://www.it.lut.fi/project/imageret/diaretdb0/

23. Doshi D, Shenoy A, Sidhpura D, Gharpure P (2016) Diabetic retinopathy detection using deep convolutional neural networks. In: 2016 International Conference on Computing, Analytics and Security Trends (CAST). IEEE, pp 261–266
24. Dridb dataset. https://www.ipg.fer.hr/ipg/resources/image_database
25. Drive dataset. https://www.isi.uu.nl/Research/Databases/DRIVE/
26. E-ophtha dataset. http://www.adcis.net/en/Download-Third-Party/E-Ophtha.html
27. E. T. D. R. S. R. Group et al (1991) Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified airlie house classification: Etdrs report number 10. Ophthalmology 98(5):786–806
28. Fan Z, Lu J, Wei C, Huang H, Cai X, Chen X (2018) A hierarchical image matting model for blood vessel segmentation in fundus images. IEEE Trans Image Process 28(5):2367–2377
29. Feng S, Zhuo Z, Pan D, Tian Q (2020) Ccnet: A cross-connected convolutional network for segmenting retinal vessels using multi-scale features. Neurocomputing 392:268–276
30. Fu H, Cheng J, Xu Y, Wong DWK, Liu J, Cao X (2018) Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE Trans Med Imaging 37(7):1597–1605
31. Gargeya R, Leng T (2017) Automated identification of diabetic retinopathy using deep learning. Ophthalmology 124(7):962–969
32. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial networks. arXiv preprint arXiv:1406.2661
33. Graham B (2015) Kaggle diabetic retinopathy detection competition report. University of Warwick
34. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J et al (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316(22):2402–2410
35. Hasan Md K et al (2021) DRNet: Segmentation and localization of optic disc and fovea from diabetic retinopathy image. Artif Intell Med 111:102001
36. Hatanaka Y, Ogohara K, Sunayama W, Miyashita M, Muramatsu C, Fujita H (2018) Automatic microaneurysms detection on retinal images using deep convolution neural network. In: 2018 International Workshop on Advanced Image Technology (IWAIT). IEEE, pp 1–2
37. He W et al (2021) Incremental learning for exudate and hemorrhage segmentation on fundus images. Inform Fusion 73:157–164
38. Hernandez-Matas C, Zabulis X, Triantafyllou A, Anyfanti P, Douma S, Argyros AA (2017) Fire: fundus image registration dataset. Model Artif Intell Ophthalmol 1(4):16–28
39. Hsu W, Pallawala P, Lee ML, Eong K-GA (2001) The role of domain knowledge in the detection of retinal hard exudates. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, vol 2. IEEE, pp II–II
40. Hussain S et al (2022) DilUnet: a U-net based architecture for blood vessels segmentation. Comput Methods Programs Biomed 218:106732
41. Indian diabetic retinopathy image dataset idrid, nov (2019) https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid. Accessed 14 June 2021
42. Isola P, Zhu J-Y, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1125–1134
43. Jin Q, Meng Z, Pham TD, Chen Q, Wei L, Su R (2019) Dunet: a deformable network for retinal vessel segmentation. Knowl Based Syst 178:149–162
44. Kaggle dataset. https://www.kaggle.com/c/diabetic-retinopathy-detection/data
45. Kaur J, Mittal D (2017) A generalized method for the detection of vascular structure in pathological retinal images. Biocybern Biomed Eng 37(1):184–200
46. Khojasteh P, Júnior LAP, Carvalho T, Rezende E, Aliahmad B, Papa JP, Kumar DK (2019) Exudate detection in fundus images using deeply-learnable features. Comput Biol Med 104:62–69
47. Kolb H (1995) Simple anatomy of the retina [internet]. Webvision: the organization of the retina and visual system. University of Utah Health Sciences Center, Salt Lake City
48. Lam C, Yi D, Guo M, Lindsey T (2018) Automated detection of diabetic retinopathy using deep learning. AMIA Summits Transl Sci Proc 2018:147
49. Lecouat B, Chang K, Foo C-S, Unnikrishnan B, Brown JM, Zenati H, Beers A, Chandrasekhar V, Kalpathy-Cramer J, Krishnaswamy P (2018) Semi-supervised deep learning for abnormality classification in retinal images. arXiv preprint arXiv:1812.07832
50. Leopold HA, Orchard J, Zelek JS, Lakshminarayanan V (2019) Pixelbnn: augmenting the pixelcnn with batch normalization and the presentation of a fast architecture for retinal vessel segmentation. J Imaging 5(2):26
51. Li F, Liu Z, Chen H, Jiang M, Zhang X, Wu Z (2019) Automatic detection of diabetic retinopathy in retinal fundus photographs based on deep learning algorithm. Translational Vis Sci Technol 8(6):4–4

52. Li T, Gao Y, Wang K, Guo S, Liu H, Kang H (2019) Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. Inf Sci 501:511–522
53. Liao Y et al (2021) Microaneurysm detection in fundus images based on a novel end-to-end convolutional neural network. Biocybern Biomed Eng 41(2):589–604
54. Lokuarachchi D, Gunarathna K, Muthumal L, Gamage T (2019) Automated detection of exudates in retinal images. In: 2019 IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA). IEEE, pp 43–47
55. Messidor dataset. http://www.adcis.net/en/Download-Third-Party/Messidor.htmlindexen
56. Naqvi SAG, Zafar MF, ul Haq I (2015) Referral system for hard exudates in eye fundus. Comput Biol Med 64:217–235
57. Nguyen QH, Muthuraman R, Singh L, Sen G, Tran AC, Nguyen BP, Chua M (2020) Diabetic retinopathy detection using deep learning. In: Proceedings of the 4th International conference on machine learning and soft computing, pp 103–107
58. Nie D, Trullo R, Lian J, Petitjean C, Ruan S, Wang Q, Shen D (2017) Medical image synthesis with context-aware generative adversarial networks. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 417–425
59. Niu Y, Gu L, Lu F, Lv F, Wang Z, Sato I, Zhang Z, Xiao Y, Dai X, Cheng T (2019) Pathological evidence exploration in deep retinal image diagnosis. In: Proceedings of the AAAI conference on artificial intelligence, vol 33, pp 1093–1101
60. Noh KJ, Park SJ, Lee S (2019) Scale-space approximated convolutional neural networks for retinal vessel segmentation. Comput Methods Programs Biomed 178:237–246
61. Pekala M, Joshi N, Liu TA, Bressler NM, DeBuc DC, Burlina P (2019) Deep learning based retinal oct segmentation. Comput Biol Med 114:103445
62. Prasad DK, Vibha L, Venugopal K (2015) Early detection of diabetic retinopathy from digital retinal fundus images. In: 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS). IEEE, pp 240–245
63. Pundikal M, Holi MS (2021) Microaneurysms detection using grey wolf optimizer and modified K-nearest neighbor for early diagnosis of diabetic retinopathy
64. Purwanithami HA, Sari CA, Rachmawanto EH (2020) Hemorrhage diabetic retinopathy detection based on fundus image using neural network and FCM segmentation. 2020 International Seminar on Application for Technology of Information and Communication (iSemantic). IEEE
65. Qiao L, Zhu Y, Zhou H (2020) Diabetic retinopathy detection using prognosis of microaneurysm and early diagnosis system for non-proliferative diabetic retinopathy based on deep learning algorithms. IEEE Access 8:104292–104302
66. Qomariah DUN, Tjandrasa H, Alam BR (2021) Hemorrhage segmentation in retinal images using modified FCN-8. In: 2021 Fourth International Conference on Vocational Education and Electrical Engineering (ICVEE). IEEE
67. Quellec G, Charrière K, Boudi Y, Cochener B, Lamard M (2017) Deep image mining for diabetic retinopathy screening. Med Image Anal 39:178–193
68. Qureshi I, Ma J, Shaheed K (2019) A hybrid proposed fundus image enhancement framework for diabetic retinopathy. Algorithms 12(1):14
69. Rokade PM, Manza RR (2015) Automatic detection of hard exudates in retinal images using haar wavelet transform. eye 4(5):402–410
70. Sahu S, Singh AK, Ghrera S, Elhoseny M et al (2019) An approach for de-noising and contrast enhancement of retinal fundus image using clahe. Opt Laser Technol 110:87–98
71. Sampath Dakshina Murthy A, Karthikeyan T, Vinoth Kanna R (2021) Gait-based person fall prediction using deep learning approach. Soft Comput. https://doi.org/10.1007/s00500-021-06125-1
72. Schlegl T, Seeböck P, Waldstein SM, Schmidt-Erfurth U, Langs G (2017) Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: International conference on information processing in medical imaging. Springer, pp 146–157
73. Scotland GS, McNamee P, Fleming AD, Goatman KA, Philip S, Prescott GJ, Sharp PF, Williams GJ, Wykes W, Leese GP et al (2010) Costs and consequences of automated algorithms versus manual grading for the detection of referable diabetic retinopathy. Br J Ophthalmol 94(6):712–719
74. Selcuk T, Alkan A (2019) Detection of microaneurysms using ant colony algorithm in the early diagnosis of diabetic retinopathy. Med Hypotheses 129:109242
75. Shah A, Lynch S, Niemeijer M, Amelon R, Clarida W, Folk J, Russell S, Wu X, Abràmoff MD (2018) Susceptibility to misdiagnosis of adversarial images by deep learning based retinal image analysis algorithms. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, pp 1454–1457

76. Shankar K, Sait ARW, Gupta D, Lakshmanaprabu S, Khanna A, Pandey HM (2020) Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model. Pattern Recognit Lett 133:210–216
77. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. J Big Data 6(1):1–48
78. Sivaswamy J et al (2018) Retinal image synthesis for cad development. In: International conference image analysis and recognition. Springer, pp 613–621
79. Son J, Park SJ, Jung K-H (2019) Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks. J Digit Imaging 32(3):499–512
80. Soomro TA, Afifi AJ, Gao J, Hellwich O, Zheng L, Paul M (2019) Strided fully convolutional neural network for boosting the sensitivity of retinal blood vessels segmentation. Expert Syst Appl 134:36–52
81. Stare dataset. http://www.cecas.clemson.edu/~ahoover/stare/
82. Stolte S, Fang R (2020) A survey on medical image analysis in diabetic retinopathy. Med Image Anal 64: 101742
83. 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
84. Tan JH, Fujita H, Sivaprasad S, Bhandary SV, Rao AK, Chua KC, Acharya UR (2017) Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network. Inf Sci 420:66–76
85. Taylor R, Batey D (2012) Handbook of retinal screening in diabetes: diagnosis and management. Wiley
86. Tchinda B, Saha et al (2021) Retinal blood vessels segmentation using classical edge detection filters and the neural network. Inf Med Unlocked 23:100521
87. Ting DSW, Cheung GCM, Wong TY (2016) Diabetic retinopathy: global prevalence, major risk factors, screening practices and public health challenges: a review. Clin Exp Ophthalmol 44(4):260–277
88. Uribe-Valencia LJ, Martínez-Carballido JF (2019) Automated optic disc region location from fundus images: Using local multi-level thresholding, best channel selection, and an intensity profile model. Biomed Signal Process Control 51:148–161
89. Wang S, Tang HL, Hu Y, Sanei S, Saleh GM, Peto T et al (2016) Localizing microaneurysms in fundus images through singular spectrum analysis. IEEE Trans Biomed Eng 64(5):990–1002
90. Wang Z, Yin Y, Shi J, Fang W, Li H, Wang X (2017) Zoom-in-net: deep mining lesions for diabetic retinopathy detection. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 267–275
91. Wang X, Lu Y, Wang Y, Chen W-B (2018) Diabetic retinopathy stage classification using convolutional neural networks. In: 2018 IEEE International Conference on Information Reuse and Integration (IRI). IEEE, pp 465–471
92. Wang L, Liu H, Lu Y, Chen H, Zhang J, Pu J (2019) A coarse-to-fine deep learning framework for optic disc segmentation in fundus images. Biomed Signal Process Control 51:82–89
93. Wang X, Jiang X, Ren J (2019) Blood vessel segmentation from fundus image by a cascade classification framework. Pattern Recogn 88:331–341
94. Wang Z et al (2022) A new detection model of microaneurysms based on improved FC-DenseNet. Sci Rep 12(1):1–9
95. Washington RE, Orchard TJ, Arena VC, LaPorte RE, Secrest AM, Tull ES (2014) All-cause mortality in a population-based type 1 diabetes cohort in the us virgin islands. Diabetes Res Clin Pract 103(3):504–509
96. Welander P, Karlsson S, Eklund A (2018) Generative adversarial networks for image-to-image translation on multi-contrast mr images-a comparison of cyclegan and unit. arXiv preprint arXiv:1806.07777
97. Wisaeng K, Sa-Ngiamvibool W (2019) Exudates detection using morphology mean shift algorithm in retinal images. IEEE Access 7:11946–11958
98. Wu Y, Hu Z (2019) Recognition of diabetic retinopathy basedon transfer learning. In: 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). IEEE, pp 398–401
99. Wu J, Zhang S, Xiao Z, Zhang F, Geng L, Lou S, Liu M (2019) Hemorrhage detection in fundus image based on 2d gaussian fitting and human visual characteristics. Opt Laser Technol 110:69–77
100. Xia H et al (2021) A multi-scale segmentation-to-classification network for tiny microaneurysm detection in fundus images. Knowl Based Syst 226:107140
101. Xu K, Feng D, Mi H (2017) Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image. Molecules 22(12):2054
102. Yi X, Walia E, Babyn P (2019) Generative adversarial network in medical imaging: a review. Med Image Anal 58:101552
103. Yu Z, Xiang Q, Meng J, Kou C, Ren Q, Lu Y (2019) Retinal image synthesis from multiple-landmarks input with generative adversarial networks. Biomed Eng Online 18(1):1–15

104. Yun WL, Acharya UR, Venkatesh YV, Chee C, Min LC, Ng EYK (2008) Identification of different stages of diabetic retinopathy using retinal optical images. Inf Sci 178(1):106–121
105. Zhao G, Fu D, Yang T (2022) A deep learning method for microaneurysms segmentation in fundus images. In: Jia Y, Zhang W, Fu Y, Yu Z, Zheng S (eds) Proceedings of 2021 Chinese intelligent systems conference. Lecture notes in electrical engineering, vol 805. Springer, Singapore. https://doi.org/10.1007/978-981-16-6320-8_79
106. Zhou W, Wu C, Chen D, Yi Y, Du W (2017) Automatic microaneurysm detection using the sparse principal component analysis-based unsupervised classification method. IEEE Access 5:2563–2572
107. Zhou K, Gao S, Cheng J, Gu Z, Fu H, Tu Z, Yang J, Zhao Y, Liu J (2020) Sparse-gan: sparsity-constrained generative adversarial network for anomaly detection in retinal oct image. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 1227–1231
108. Zhou Y, Chen Z, Shen H, Zheng X, Zhao R, Duan X (2021) A refined equilibrium generative adversarial network for retinal vessel segmentation. Neurocomputing 437:118–130

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

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

Archives of Computational Methods in Engineering
https://doi.org/10.1007/s11831-022-09823-7

REVIEW ARTICLE

A Systematic Review on Artificial Intelligence-Based Techniques for Diagnosis of Cardiovascular Arrhythmia Diseases: Challenges and Opportunities

Shikha Singhal[1] · **Manjeet Kumar**[1]

Received: 22 April 2022 / Accepted: 4 September 2022
© The Author(s) under exclusive licence to International Center for Numerical Methods in Engineering (CIMNE) 2022

Abstract
Cardiovascular health-related problem is a rapidly increasing integrated field concerning the processing and fetching the information from cardiovascular systems for early detection and treatment of cardiovascular diseases. Artificial Intelligence (AI) techniques, especially machine and deep learning techniques are more impactful and powerful tools for upgrading the capabilities of an application, and they have been applied to medical data for analysis and disease detection purposes. This paper, represents a comprehensive view of AI-based computational modeling with the abilities of powerful AI techniques that can play a crucial role in developing smart and enhanced systems in a real-world application. The paper outlines the broad overview of AI-based modeling that can be utilized in various application domains. An electrocardiogram (ECG) plays a major role in biomedical applications to record the heartbeat activity. Regular monitoring of ECG through wearable devices like the band, watches, etc. can be done for early detection of cardiovascular diseases. The competency of each method discussed is related to ECG classification approaches that have been compared in terms of some parameters like accuracy, sensitivity, specificity, positive predictivity, and F-score. The noise affects the ECG signal which may deteriorate the features of the respective signal that leading to improper treatment. De-noising has been done by pre-processing of the signal, which enables the prediction of the heart condition. After detecting the positions of the P wave, QRS wave, and T wave, feature extraction has been done. The efficiency of ECG classification with different computational methods was evaluated with proposed algorithms by using different databases like MIT-BIH, PTB, and MIT-BIH Atrial Fibrillation test. It has been observed from the literature that Convolutional Neural Networks (CNN) are best suited for the classification and detection of arrhythmia. The challenges of existing techniques to analyze the ECG signal for the classification and detection of arrhythmia are summarized.

1 Introduction

In the last few decades, cardiovascular diseases (CVDs) are the major cause of death and fatality in the world [1]. For processing the encrypted signals, more attention has been given to the development of various tools in the last few years [2]. Approximately, 17.9 million people died from CVDs every year globally which represents 35% of total global deaths [3]. Early detection and diagnosis of CVDs are one of the leading research areas in today's world. The regular monitoring of heart rate is the key factor to analyse the cardiovascular health status of an individual. An electrocardiogram (ECG) has been a significant and unique tool in investigating and recognizing cardiovascular diseases [4]. Any deviation in the heart rate is an indication of some fundamental pathology changes which can be identified by the ECG signal. ECG signal represents the time evolution of the electrical activity of the heart which consists of electrical depolarisation-repolarisation patterns of the heart [5]. By monitoring the heart rate or rhythm, abnormalities in the functioning of the heart can be identified. Body sensor networks are helpful to diagnose the irregularity in the activity of the heart [6]. In recent years, the analysis of ECG signals has drawn much attention to the automatic detection and diagnosis of heart disease [7]. There are several advantages of ECG technology for heart rate monitoring, but the

✉ Manjeet Kumar
manjeetchhillar@gmail.com

Shikha Singhal
ggshikha12@gmail.com

[1] Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India

Published online: 30 September 2022

major drawback is the noise introduced during ECG signal acquisitions like baseline wander, power line interference, EMG noise, and motion artifacts [8] which degrades the performance of the signal.

1.1 Basic Principle of ECG Signal

ECG signal measures the electrical activity of the heart and the deviation from the normal one through which the disease can be detected. ECG comprises the depolarisation-repolarisation of cells after the rest membrane potential (RMP) reaches the threshold. When sodium ions (Na^+) strike the sinus-atrial (SA) node, the rest membrane potential reaches the threshold and loses the electronegativity after which depolarization occurs. After the completion of depolarization, potassium ions (K^+) leave the cell area, and calcium (Ca^+) ions reach the cell area. The process continues and after a few cycles, it loses all of its electropositivity through potassium (K^+) ions and the cell attains electronegativity again. This process represents repolarization [9].

During the depolarization process, all the cells get depolarised in the atrium and a total vector of the electromagnetic force is directed towards the left and downward near the ventricle. As the atria myocardium is thin and does not comprise any specialized conduction system, hence, the generation of the vector is small and directed with moderate velocity towards the ventricle. Now, depolarization occurs in the ventricular area through the atrial ventricular (AV) node which is a slow conductor. Once it enters the ventricular area through the AV node, conduction takes place very fast due to Purkinje fibers and depolarization occurs. After depolarization of the ventricular area, repolarisation occurs in the next step. This makes the complete process of depolarisation and repolarisation. The ECG waveform along with different segments and intervals is shown in Fig. 1.

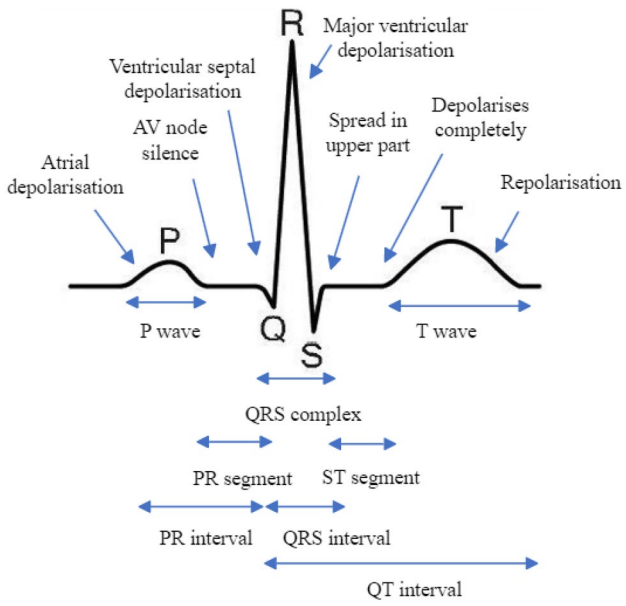


Fig. 1 Representation of different segments and intervals of ECG signal

Table 1 Summarizes the amplitude, frequency range, and duration of different ECG waves. The duration of different intervals of ECG waveform is reported in Table 2.

ECG leads are split into bipolar limb leads and unipolar leads. The twelve leads of the ECG recorder represent twelve electrical views of the heart from twelve different angles. The leads of the ECG recorder are depicted in Fig. 2.

The conventional 12-lead procedure involves connecting ten electrodes to the body: one to each limb and six across the chest. There are six limb leads and six chest leads [9]. The six limb leads look at the heart in a vertical plane and are obtained from three electrodes attached to the right arm, left arm, and left leg. The electrode on the right leg is an earth electrode. The measurement of the voltage requires two poles: negative & positive. The negative pole is used as zero references. Limb leads I, II, and III are bipolar as they measure the electric potential between two of the three limb electrodes. Leads aVR, aVL, and aVF are unipolar leads. They use one limb electrode as the positive pole & take the average of inputs from the other two as the zero references. Hence, aVR looks at the upper right side of the heart; aVL looks at the upper left side of the heart, and aVF looks at the inferior wall of the heart. From twelve leads, the magnitude of the heart's electrical potential can be measured and recorded over some time. The details about lead placement are reported in Table 3.

The positioning of different bipolar, unipolar, and chest leads is shown in Fig. 3. Chest leads are unipolar leads and

Table 1 Amplitude and frequency range with the duration of different ECG waves

Features	Amplitude (mV)	Frequency (Hz)	Duration (ms)
P-wave	0.20–0.25	8–10	0.08–0.10
R-wave	1.40–1.60	–	0.08–0.12
Q-wave	0.20–0.40	–	0.09–0.1
T-wave	0.10–0.50	5–8	0.12–0.16
QRS-complex	–	10–40	0.09–0.10

Table 2 Different intervals of ECG signal

Features	Duration (ms)
PR-interval	0.12–0.22
QT-interval	0.35–0.44
ST-interval	0.05–0.15
QRS-interval	0.09–0.10
RR-interval	0.6–1.2

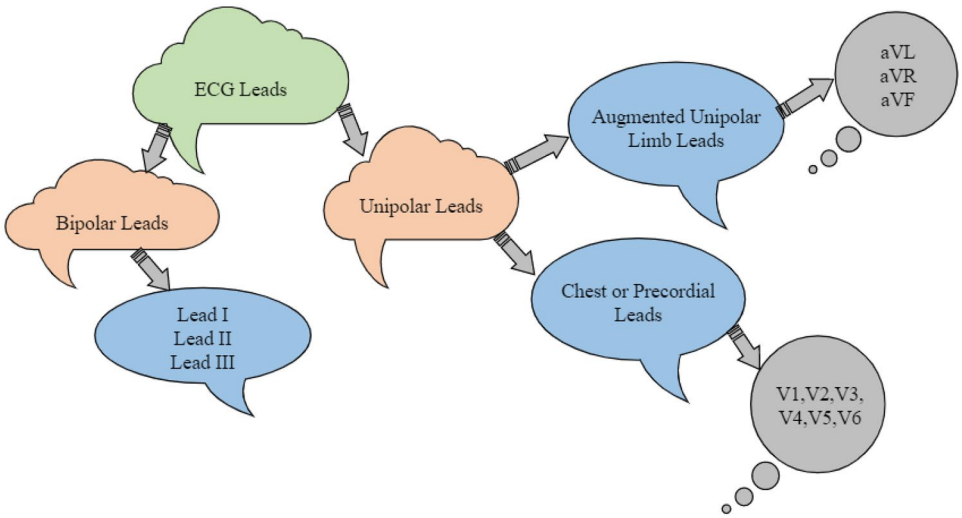


Fig. 2 Representation of different ECG leads of 12-lead ECG recorder

Table 3 Summary of lead placement for 12-lead ECG recorder [9]

Types of leads	Leads	Placement of leads
Bipolar leads	Lead I	LA–RA (left arm-right arm) (between the left arm and right arm)
	Lead II	LL–RA (left leg-right arm) (between left leg and right arm)
	Lead III	LL–LA (left leg-left arm) (between left leg and left arm)
Unipolar leads	Lead aVR(right arm)	RA (+) to [LA & LF] (−) (Rightward)
	Lead aVL (left arm)	LA (+) to [RA & LF] (−) (Leftward)
	Lead aVF (left leg)	LF (+) to [RA & LA] (−) (Inferior)
	Unipolar chest leads or precordial leads	V1,V2,V3,V4,V5,V6

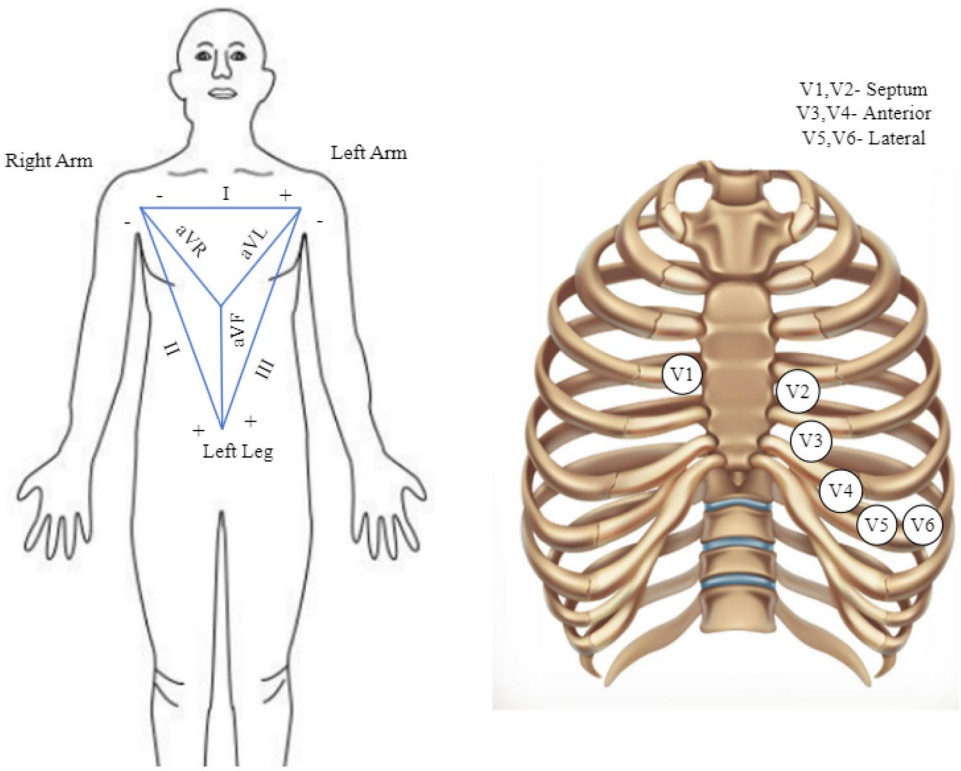


Fig. 3 Positions for the placement of unipolar limb leads, bipolar limb leads, and precordial leads

are used to determine the potential difference between the positive electrode and virtual electrode. V1 and V2 are oriented over the right side of the heart but they are sensitive to septal electrical activity of the heart. Functionally, V1 and V2 are septal leads. V3 and V4 are oriented over the septum and are sensitive to anterior myocardial infarction. V5 and V6 are positioned at the left that are sensitive to major ventricular depolarisation. If infarction occurs in (V1, V2, V3, V4) and (V3, V4, V5, V6) are referred to as anteroseptal leads and anterolateral leads, respectively [9].

1.2 Heart Rate Estimation Using ECG

The average number of times a heart beats per minute (bpm) represents the heart rate. The deviation in heart rate from the normal range shows cardiac arrhythmia [10]. Arrhythmia is a condition where heartbeats show an irregular pattern. The normal heart rate is about 70 bpm. During the activity, it indicates a disorder known as bradycardia when the heart rate becomes lower than 60 bpm. During athletic activity, the instantaneous heart rate can reach up to 200 bpm which exhibits a disorder named tachycardia [11]. The classification of different types of arrhythmias based on heart rate is listed in Table 4.

Table 4 Heart rate relationship with the categorization of different types of arrhythmias [11]

Range of heart rate(bpm)	Types of arrhythmias
> 350	Atrial fibrillation
250–350	Atrial flutter
150–250	Paroxysmal tachyarrhythmia
100–150	Simple tachyarrhythmia
40–60	Mild bradyarrhythmia
20–40	Moderate bradyarrhythmia
0–20	Severe bradyarrhythmia

Arrhythmia which originates in the sinus and atria is classified into a different category. Sinus Arrhythmia is divided into physiological sinus arrhythmia, sinus tachycardia, and sinus bradycardia. Physiological sinus arrhythmia shows the inspiration and expiration phase results in different RR-interval [12]. During inspiration, RR-interval decreases as heart rate increases due to vagal inhibitory action, and RR-interval increases during expiration by vagal stimulatory action. In Sinus tachycardia overstimulation is done on the Sinoatrial (SA) node that increases the heart rate, occurs due to fever, more exercise, etc. Sinus bradycardia occurs in athletes, patients with hyperthermia, jaundice, etc. which slow down the activity of an SA node. SA node initially works well but later it slows down which is termed tachybradysyndrome or sick sinus syndrome [12].

Whereas, atrial tachyarrhythmia is split into atrial tachycardia, atrial flutter, and atrial fibrillation. In Atrial tachycardia, multiple P-waves followed by one QRS-complex show ventricular activity. A very rapid and fast rhythm appears in atrial flutter. In atrial fibrillation, directed vectors are not synchronized in the atrium due to which there is no electrical vector in the atrial area. Different stages of ECG classification are shown in Fig. 5.

Major steps involved in the overall ECG classification process can be represented in Fig. 4.

The remaining paper is organized as follows: The analysis and classification of ECG signals are described in Sect. 2. Section 3 outlines the database studied in previous papers and contributes a comprehensive specific data

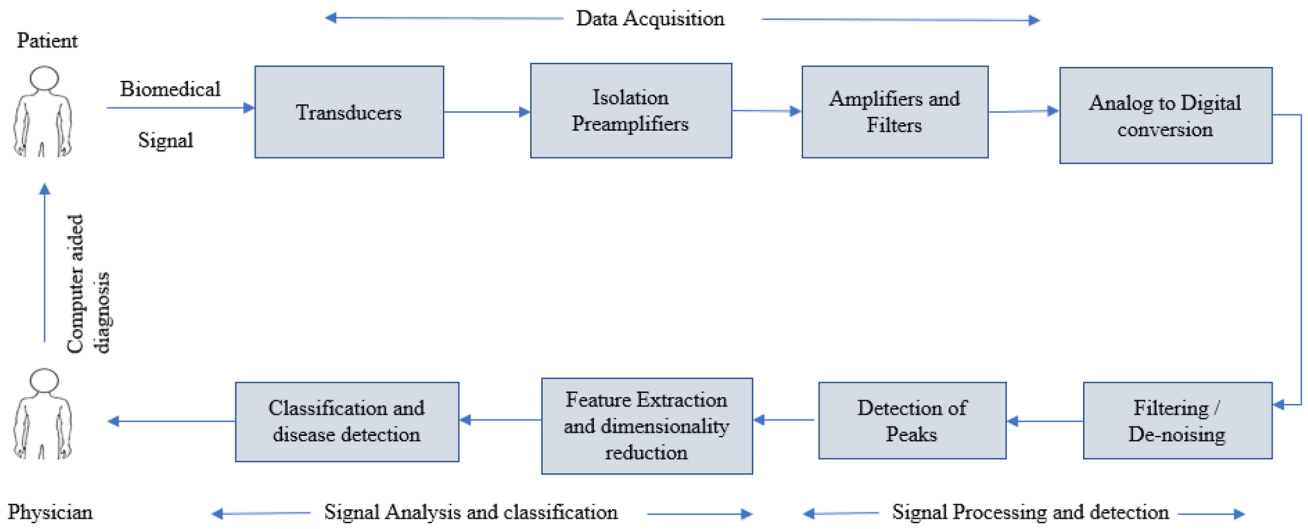


Fig. 4 Computer aided-diagnosis based on Biomedical signal processing and classification

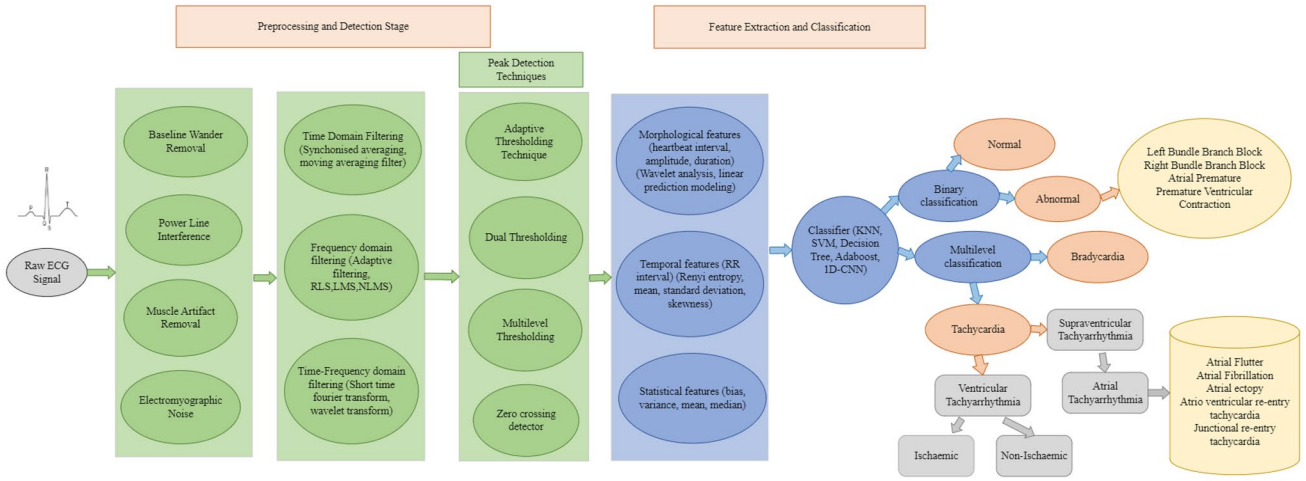


Fig. 5 Block diagram representation of different stages in the classification of ECG signal

characterization for the proposed heartbeat classification system. Section 4 introduces the methodology which has been employed by different researchers along with the results obtained by different methods. The challenges and limitations of the available techniques are summarized and possibilities for further improvement are discussed in Sect. 5. Finally, Sect. 6 concludes the findings of different classification methods to detect arrhythmia.

2 Basic Process of ECG Signal Analysis and Classification

ECG monitoring describes a set of common processes as data acquisition, pre-processing, feature extraction, and classification as shown in Fig. 5. Data acquisition with wired or wireless sensors measures and records the activity of the heart. A high pass sigma-delta modulator is used for acquisition and digitization [6]. ECG signal is required to be pre-processed after data acquisition.

2.1 Pre-processing

It is essential to remove noise from the ECG signals. In the acquisition stage, different frequency components are used that may produce interference in the signal recording process and may add noise to the ECG signal. This unwanted noise in the signal may corrupt the original information and leads to erroneous data. Filtering methods are implemented to remove the noise from the corrupted ECG signal. Based on the type of noise present in the signal, an appropriate filter is to be selected. Baseline wander, Power line interference (PLI), Motion artifact, and Electromyographic (EMG) noise are the main noises that have to be suppressed for further processing. Baseline wander is a low-frequency noise in the range of 0.15–0.3 Hz caused by respiration that creates a hurdle in the ECG signal. With the increase in breathing rate, baseline wandering increases. Low pass filter, wavelet transformation is implemented to remove the baseline wandering [1, 13, 14] for the signal to be processed. Motion artifact occurs due to the movement of the electrode over the skin resulting from the movement of the patient. EMG noise occurs due to the shrinking of muscles which reflects the potential in the range of millivolts. PLI is due to the interference between a power line and the ECG recording equipment. For the removal of PLI, data are passed through the median filters which are used to remove the QRS complexes and T wave. Finally, a notch filter centered at 60 Hz (or 50 Hz) is used to remove the interference [3]. A low pass Butterworth filter can be used as a powerline interference canceller [13]. A wavelet filter is used to remove the Gaussian white noise, wavelet transform is applied to ECG signals for reducing the noise factor [5]. De-noising of ECG signal is performed by non-linear Bayesian filtering [15]. The main motive of pre-processing step is to reduce the noise level from the corrupted ECG signal to enhance the accuracy of the system. The detail of different pre-processing techniques available in the literature is presented in [8]. About [16], a comparison has been done in several techniques to denoise the corrupted signal. Time-domain, frequency-domain, time–frequency domain (wavelet transform), and others proposed in the last few decades are used as denoising techniques.

2.2 Feature Extraction

2.2.1 Feature Sets

For any given task, an optimal set of features exists among a large number of features from the feature set [1]. Features

are categorized into morphological features, temporal features, time-domain features, frequency-domain features, and non-linear features.

(1) *Morphological features* Based on the position and amplitude of the P wave, Q wave, R wave, S wave, and T wave that calculates the depths and specifies the area, and power, of any signal. In ECG signal, P, Q, R, S, T peak points, QRS-complex duration, PR-interval, QT-interval, and ST-interval are the features extracted as morphological features.
(2) *Temporal features* These are the most relevant heart rate features based on the length of intervals (RR-interval) including the ECG wave acquired from time-domain signals [1, 10].
(3) *Time-domain features* Statistical features completed the set of features including the maximum and minimum value of RR interval, histogram, variance, mean, standard deviation, and skewness.
(4) *Frequency-domain features* Includes the wavelet transform coefficient, normalized low and high-frequency power, based on the ECG signal with a window. Spectral characteristics of the window are calculated as center frequency, and spectral frequency [10].
(5) *Non-linear features* It has been calculated by a non-linear method like fuzzy entropy, sample entropy, etc. The detailed feature of the ECG signal is summarised in Table 5.

This step is used to extract useful information from ECG signals for disease classification. All the characteristic waves are identified with their intervals in the feature extraction stage. It has been observed from the previous studies that the wavelet approach is the most favourable feature extraction approach. For the multiresolution analysis of ECG, discrete wavelet transform is employed [14] to attain more accurate results. Some distinctive noise-specific features like short-term zero crossing for detection and classification are used [17]. In ECG diagnosis, R-peaks are detected by frequency-domain analysis, and time-domain analysis is used to detect other characteristic waves [18]. The most appropriate technique for detection purposes is the thresholding technique in which pre-processed signal is compared with the adaptive threshold to detect the QRS-complex. Some dimensionality reduction techniques like principal component analysis (PCA) and independent component analysis (ICA) are implemented to extract features from the ECG signal [19]. With these methods, the important part of information can be preserved by projecting to a low-dimensional space. Four temporal features are computed for the classification of each beat of Pre RR-interval, Post RR-interval, Local average RR-interval, and Global average RR-interval.

2.3 Classification

For the detection of cardiac arrhythmia, the classification of ECG signals is necessary and it is one of the major challenges that reflect more efforts in this field [20]. Neural networks are used for the classification of ECG signals. It takes the extracted features as an input and the output is the desired result. It works as a computational algorithm based on the biological neural network. ECG Classification process consists of several classifiers which are widely used to improve the accuracy of the detection process. It has been observed from previous studies that various classifiers are used to classify the ECG signal in a modified way. The decision tree [21], linear support vector machine [22], Adaboost, SVM, ANN [23], 2-D

Table 5 Summary of the detailed features of ECG signal

Type of feature	Features
Morphological feature (based on amplitude, power, area)	QRS morphology P and T wave morphology PR-interval QT-interval ST-interval
Temporal feature (based on length of the interval)	Pre RR-interval Post RR-interval Average RR-interval P-wave duration QRS interval T-wave duration
Time-domain feature (based on histogram, mean, variance, standard deviation)	Mean of RR-interval Root mean square value of the difference between RR interval
Frequency-domain feature (based on ECG signal with window)	Spectral parameters
Non-linear features	Fuzzy entropy, sample entropy

CNN [24], SVM-RBF [25], DNN, Patient independent 1-D classifier [26], Multi-perspective CNN (MPCNN) [27], 1-D CNN, Deep RNN [28] are used to classify the ECG signal with improved accuracy and other specifications.

3 Database

The ECG databases play a supreme role in the development and enhancement of algorithms that are helpful in the classification and detection process. For the analysis of ECG signals, several databases are publicly available to scrutinize the proposed algorithms. It is a platform through which the required model can result in multiple effective ways. The database increases the opportunity to produce the outcome in many different forms which will be helpful in real-world domain. The different database has been used by various researchers in the past few years. Details of some important databases are available below.

(a) Physiobank database [29] is an important database that consists of a well-characterized biomedical signal that represents 200 ECG signal recordings out of which 160 are to be treated as training data and the other 40 are treated as test data. It is used to characterize the data of patients with health implications such as myocardial infarction, sudden cardiac death, etc.

(b) Massachusetts Institute of Technology-Beth Israel Hospital Arrhythmia database (MIT-BIH) [30] is the popular database used for arrhythmia detection which consists of (48-half hours each) ambulatory ECG recording. MIT-BIH dataset consists of 23 records where (inpatients-60%, outpatients-40%), and the rest 25 records show an uncommon but clinically significant arrhythmia. This includes 25 men aged from 32 to 89 years, and 22 women aged from 23 to 89 years. The sampling rate is 360 samples per second per channel is used with a resolution of 11 bits. The publication of the MIT-BIH Arrhythmia database can be regarded as a major achievement in the development of arrhythmia detectors that enhances the performance of the system.

(c) MIT-BIH Normal Sinus Rhythm database [31] also plays a major role that includes 18 long-term ECG recordings of subjects and the dataset includes 5 men, aged 26 to 45 and 13 women aged 20 to 50. The sampling frequency is 128 Hz.

(d) MIT-BIH Atrial Fibrillation Database [32] is an important database that includes 25 long-term ECG recordings with atrial fibrillation. The duration is of 10 h each for an individual recording. Each signal is sampled at 250 samples per second with 12-bit resolution over a range of ±10mV.

(e) Physikalisch-Technische Bundesanstalt diagnostic 12-lead ECG database (PTBDB) [33] also plays a major role in the detection process of arrhythmia that consists of 549 records from 290 subjects from which 52 are considered healthy records. Each signal is sampled at 1000 samples per second with a resolution of 16 bits over a range of ± 16.384 mV. A detailed summary of 268 subjects is available, rest detailed summary of 22 subjects is not available.

(f) The Fantasia database [34] shows the effective development of detectors which consists of 40 ECG recordings, 20 records from the young generation aged 21 to 34 years, and 20 from elderly people aged 68 to 85 years. The duration of all the recordings is 120 min and the sampling rate is 250 Hz.

(g) ECG ID [35] database includes 310 recordings taken from 90 subjects. 44 men and 46 women from various fields are taken for the recordings. The sampling frequency of the ECG signal is 500 Hz with a 12-bit resolution over a range of ±10mV.

Table 6 summarizes all the publicly available databases for ECG classification and arrhythmia detection used in the available literature.

4 Literature Review Based on ECG Arrhythmia Detection and Classification

For the accurate detection purpose of ECG signal, preprocessing is the primary and crucial step that is to be implemented. Classification is required to detect the presence of an arrhythmia in the required signal. Different machine learning and deep learning algorithms have been implemented for the proper detection of arrhythmia. The literature review summarizes the work related to preprocessing, peak detection, and arrhythmia detection using different classification techniques.

4.1 Parameters to Evaluate the Performance of an ECG Algorithm

Different parameters are used to evaluate the ECG algorithm proposed by various researchers. Evaluation parameters provide a mechanism to compare the various proposed algorithms. The performance of detection is estimated by calculating the parameters like Accuracy (*Acc*), Sensitivity (*Sen*), Specificity (*Spe*), Positive predictivity (*ppv*), F-score value, and the Error rate (*Er*) of the system as defined in Eqs. (1)–(6).

$$Accuracy(Acc) = \frac{(True\ positive + True\ neagtive)}{(True\ positive + True\ negative + False\ positive + False\ negative)} \quad (1)$$

Table 6 Summary of available ECG arrhythmia detection databases

Database	Method/algorithm	Number of subject/data	Sampling Frequency/Range
Physiobank database [29]	NN (decision tree) [2]	ECG signal records = 200 Training set = 160; Test set = 40	NR
MIT-BIH arrhythmia database [30]	Wavelet transform(Sequential forward floating search (SFFS) algorithm) [1], Time based Integrate and fire (IF) Sampler(simple binning approach) [3], Reservoir computing, logistic regression(Ring configuration) [4], ANN(PSO) [13], CNN(PCA) [36], Abductive interpretation (QRS clustering) [20], KNN [37], CEEMD [17], (GCNN, TDCNN) [19], ANN [38], Gaussian-Bernoulli deep belief network [22], 2D CNN [24], Cross validation(PCA and DTW) [25], Regularization (MLP) [39], K fold cross validation [26], MPCNN (SAX) [27], CNN (BLSTM & BGRU) [28], LSTM RNN (GRU) [40], LSTM (SDAE) [41], CNN(STFT) [42], SNN(STDP and RSTDP) [43], CNN(BLSTM) [44], MLPNN (ISVS) [45], Decision tree, Gradient boosted tree [46], CNN(GAN) [35], CNN (Wavelet filters) [47], KNN (AKMC) [48], Ensemble SVM [49], Residual-concatenate network (ITML) [50], CNN + DNN (DivNet) [51], CNN (FCResNet, MOWPT) [52], SNN [53], wavelet transform-based digital signal processing circuit, DBM (Template based) [54], Model driven architecture, MDA [55]	Number of records = 44 Dataset1 = 22 (training) Dataset2 = 22 (final performance) Ambulatory ECG recordings = 48 (half hour each) BIH database (inpatients-60%; outpatients-40% = 23 recordings) Clinically significant arrhythmias = 25 recordings Men = 25 (aged 32 to 89) Women = 22(aged 23 to 89)	Sampled at 360 Hz Resolution = 11 bits, Range = ± 10 mV
MIT-BIH Normal Sinus Rhythm database [31]	Multiclass learning approach (fast optimization algorithm) [56], CEEMD [17]	Number of significant arrhythmias = 18 subjects Men = 5, aged 26 to 45; Women = 13, aged 20 to 50 Training data = from hours 1 to 2 Test data = ECG signal from hour 13 to 14	Sampled at 128 Hz
MIT-BIH Atrial Fibrillation Database [32]	KNN (Burg's method, Atrial Fibrillation) [57], CNN (CAD) [58]	Atrial Fibrillation recordings = 23	NR
Physikalisch-Technische Bundesanstalt diagnostic 12-lead ECG database (PTBDB) [33]	DWT (Cross wavelet transform) [5], CEEMD [17], MLPNN (ISVS) [45], DWT (DEA) [59]	Subjects = 290; Records = 549 Healthy controls = 52; Myocardial Infarction patients = 148	Sampled at 1 kHz Resolution = 16 bit Range = ± 16.384 mV
Fantasia database [34]	CEEMD [17]	Young subjects = 20 Elderly subjects = 20 Continuous supine resting ECG signals obtained = 120 min	Sampled at 250 Hz
MIT-BIH ECG ID database signals [35]	Wavelet transform approach (ANN) [18], ANN (Peak detection algorithm) [18], CNN + RNN (BLSTM + BGRU) [28]	Total recordings = 5066 ECG recordings = 310, from 90 persons 44 men & 46 women Age = 13 to 75 years	Sampled at 500 Hz Resolution = 12 bit Nominal range = ± 10 mV

Table 6 (continued)

Database	Method/algorithm	Number of subject/data	Sampling Frequency/Range
Physionet. PhysioNet/Computing in Cardiology Challenge 2011(PCICC2011) database [47]	CEEMD [17], OCSVM [60]	Twelve-lead ECG recording = 2000	Sampled at 500 samples per second Resolution = 16 bit
MIT-BIH Supra-ventricular Arrhythmia Database (SVDB) [61]	Gaussian-Bernoulli deep belief network [22], LSTM (SDAE) [41], DCN (DenseNet) [42], GDB (Decision tree) [46]	78 two-lead recordings, 30 min long Performance evaluation = first 70 records	Sampled at 128 Hz
The MIT-BIH Noise Stress Test (NST) Dataset [62]	LSTM (SDAE) [41], CNN (CAD) [58]	ECG recordings = 12 half-hour Noise Half-hour recordings = 3	Sampled at 360 Hz
The Local hospital, courtesy of the SiWei medical company (Shanghai, China) and SiWei Remote ECG diagnostic center	Tensor learning approach (Generalized tensor rank one discriminant analysis) (GTR1DA) [7]	Total pieces of ECG data = 98,287; Kind of single or mixed diseases = 1251 Single-disease categories = 249 High quality 12-lead ECG records = 3,000 pieces	The sampling frequency is 500 Hz
ECG dataset was obtained from HCM patients and all subjects, ICD patients [63]	Cross validation [10]	HCM patients: no of patient = 221 Total ECG recordings = 754 Total heartbeats = 6488 ICD (control):no of patient = 541 Total ECG recordings = 541 Total heartbeats = 4442	NR
Incart databases [64]	SKF (Non-linear bayesian filtering) [15], DCN (DenseNet) [42], CNN (Segmentation) [65]	Incart database = 75 recordings > 175,000 annotated beats	Sampled at 257 Hz
1631 3-s ECG segments with clinical rhythm annotations [66]	Gaussian mixture model (LDA, QDA) [14]	Obtained from 298 out-of-hospital cardiac arrest patients	Sampled at 500 Hz Resolution = 16 bit ECG had a resolution of 1.031 μV per least significant bit
MIT-BIH ST Change (MIT-BIHSTC) [67]	CEEMD [17]	28-two lead ECG recordings	Sampled at 250 Hz Resolution = 12/16 bit
UCI arrhythmia dataset [68]	Decision tree approach (Genetic algorithm) [21]	452 instances; 279 attributes linear values = 206 attributes, normal group = 245 instances abnormal group = 207 instances	NR
Chinese Cardiovascular Disease Database (CCDD) dataset [69]	PPNN Model (Bayesian probability algorithm, DTW algorithm) [70]	Total records = 46,729 Total ECG labels = 270 12-lead ECG devices from different hospitals Short-term ECG records = 200,000	Sampled at 500 Hz
CinC 2017 A fib database [71]	Wave detection algorithm [72]	Total ECG recordings = 8528	Sampled at 300 Hz

SFFS sequential forward floating search, *PSO* particle swarm optimization, *ANN* artificial neural network, *CNN* convolutional neural network, *PCA* principal component analysis, *PPNN* probabilistic process neural network, *GCNN* generic convolutional neural network, *TDCNN* tuned dedicated convolutional neural network, *CEEMD* complete modified ensemble empirical mode decomposition, *DTW* dynamic time warping, *MLP* multilayer perceptron, *MPCNN* multi-perspective convolutional neural network, *SAX* symbolic aggregate approximation, *BLSTM* bidirectional long short-term memory, *GRU* gated recurrent unit, *BGRU* bidirectional gated recurrent unit, *SDAE* stacked denoising autoencoders, *STFT* short time Fourier transform, *ISVS* in-set voting scheme, *STDP* spike-timing dependent plasticity, *RSTDP* reward-modulated STDP, *MLPNN* multilayer perceptron neural network, *GAN* generative adversarial networks, *ITML* information-theoretic metric learning, *MOWPT* maximal overlap wavelet packet transform, *OCSVM* one-class support vector machine, *FCResNet* fast compression residual convolutional neural networks, *AKMC* adaptive K means clustering, *SNN* spiking neural network, *NR* not reported

$$Sensitivty(Sen) = \frac{(True\ positive)}{(True\ positive + True\ negative)} \quad (2)$$

$$\text{Specificity(Spe)} = \frac{(True\ negative)}{(True\ positive + False\ positive)} \quad (3)$$

$$\text{Positive Predictivity(ppv)} = \frac{(True\ positive)}{(True\ positive + False\ positive)} \quad (4)$$

$$F - score(Fs) = \frac{(2 \times Sensitivity \times positive\ predictivity)}{(Sensitivity + positive\ predictivity)} \quad (5)$$

$$Error\ Rate = \frac{(False\ Positive + False\ Negative)}{Total\ beat} \quad (6)$$

True positive shows an outcome where the model correctly predicts the positive class. True negative represents the observation where the model verifies the negative class. False-positive represents an outcome where the model predicts the positive observations but the value was actually negative. False-negative shows the number of observations where the model predicted a negative value but the value was actually positive. Sensitivity measures the true positive cases in overall true cases. Specificity calculates the true negative values in overall positive cases.

As it concerns with the classification of disease, we may define a confusion matrix also known as an error matrix, which describes the performance of an algorithm. The matrix is segregated into predicted and actual classes in rows and columns. It uses four kinds of results as true positive, true negative, false positive and false negative. It helps to classify the disease accurately with the test data.

Actual value	Predicted value	
Total (P + N)	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Table 6 represent the confusion matrix of predicted and actual values with test data.

Zaunseder et al. [1] revealed about ECG classification that an SFFS algorithm is applied with a new criterion function index based on linear discriminants with classifiers LDA, KNN, and MLP having an accuracy of 89%. Different classifiers like decision tree, and quadratic discriminant function (QDF) in a hybrid approach with an accuracy of 86.3% proposed by Barni et al. [2]. A time-based encoding using the binning approach with linear discriminant analysis (LDA) as a classifier having a sensitivity of 92% is used by Alvarado et al. [3]. Moran et al. [4] presented reservoir computing with logistic regression attained the accuracy and sensitivity of 98.4% and 84.85% respectively. Banerjee and Mitra [5] revealed the discrete wavelet transform-based decomposition with accuracy, sensitivity, and specificity of 97.6%, 97.3%, and 98.8%, respectively. Wavelet transform is a very effective technique that results in maximum output. Some techniques like high pass sigma-delta modulator (HPSDM) and chopper-based continuous-time amplifier (CBCTA) as biosignal processors with an accuracy of 97.25% have been recorded by Lee et al. [6]. Huang and Zhang [7] stated the generalized tensor rank one discriminant analysis (GTR1DA) with a support vector machine (SVM) classifier has an accuracy of 99.8%. Rahman et al. [10] used the cross-validation with random forest (RF), and SVM with a specificity of 0.93. Shadmand and Masoufi [13] presented the ANN and block-based neural network (BBNN) with the PSO algorithm having an accuracy of 97%. A Gaussian mixture model (GMM) with ANN, LDA, and QDA with an accuracy of 78.5% has been proposed by Eftestol et al. [14]. Oster et al. [15] stated the bayesian filtering with SVM as a classifier has a sensitivity of 99%. The fast optimization algorithm with a non-linear kernel classifier has been used by Gutta and Cheng [56]. Padmavathi and Ramakrishna [57] used Burg's method of atrial fibrillation with KSVM and attained an accuracy of 100%. Kiranyaz et al. [36] proposed CNN as a classifier with PCA and an accuracy of 97.5%. Satija et al. [17] used complete modified ensemble empirical mode decomposition (CEEMD) with temporal features having an accuracy of 97.38%. Park et al. [37] proposed the KNN algorithm with cascaded RF classifier having accuracy, sensitivity, and specificity of 97.34%, 91.23%, and 99.81% respectively.

Patro and Kumar [18] stated the cascaded filter with ANN for pattern recognition classifier accuracy of 98.14%. Li et al. [19] used the generic CNN (GCNN) and tuned dedicated CNN (TDCNN) with an accuracy of 96.89%. The clustering algorithm has been presented with a sensitivity of 88% by Teijeiro et al. [20]. Sannino and Pietro [38] stated the ANN approach with a DNN classifier has accuracy, sensitivity, and specificity of 99.68%, 99.48%, and 99.83% respectively. Ayar and Sabamoniri [21] proposed the decision tree approach with a genetic algorithm that used the hybrid model with optimal features having an accuracy of 2 class and 16 class is 86.76% and 77%, respectively. Gaussian Bernoulli deep belief network with SVM as classifier results in the accuracy of 99.5% and 99.4% in case of supraventricular and ventricular respectively with the MIT-BIH database by Sayanta et al. [22]. Filtering methods like Butterworth, bandpass filter, wavelet filter, etc. have been implemented with classifiers KNN, ANN, and Naïve Bayes having different algorithms as peak detection algorithms [23, 24, 48, 73]. Autoregressive modeling with model-driven architecture has been presented by Iqbal et al. [55]. K fold cross-validation has been done with classifiers SVM, ANN, and PCA, MLP

as an algorithm having an accuracy of 99% [25, 26, 39]. Adam optimizer with symbolic baseline-corrected approximation (SBCX) approach and MPCNN as classifier shown the accuracy of 96.4% revealed by Niu et al. [27].

In [28], Lynn et al. introduced a deep recurrent neural network (RNNs) based on Gated Recurrent Unit (GRU) in a bidirectional manner achieving an accuracy of 98.5%. Here, the 1-D CNN approach and RNN with long short-term memory (LSTM) and GRU have been used for classification. The proposed model was evaluated with ECG ID and MIT-BIH database with an accuracy of 98.6% and 98.4%, respectively. In [40], Saadatnejad et.al presented an architecture consisting of wavelet transform for feature extraction (wavelet and classical features) and multiple LSTM RNNs for classification purposes. It classifies the heartbeat into ventricular ectopic beat (VEB) and supraventricular ectopic beat (SVEB) with an accuracy of 99.2% and 98.3%, respectively. In [41], Wang et.al presented the stacked denoising autoencoders (SDAE) that learned the semantic encoding of heartbeats which helps in the reduction of noise level and correspondingly increases the classification efficiency. Bidirectional LSTM was employed as a classifier that worked on temporal features. Three databases MIT-BIH arrhythmia database, MIT-BIH Supraventricular Arrhythmia database (SVDB), and MIT-BIH Noise Stress Test (NSTDB) have been used and attained an overall accuracy, specificity, and sensitivity of 98.85%, 99.86%, and 96.32%, respectively. In [42], Rahhal et.al presented an architecture based on dense convolutional networks (DCN) for the ECG signal classification. This architecture consists of a generative and discriminative module that one employs the conversion of one-dimensional ECG input to the two-dimensional image so that image can be easily applied to the convolutional neural network and the other employs the classification of the signal. To handle the imbalanced data, introduced the idea to use focal loss that reduces the loss assigned to ECG beats. MIT-BIH Arrhythmia database, INCART database, and MIT-BIH SVDB databases have been used for evaluating the method with an overall accuracy of SVEB is 97% and VEB is 74% respectively. In [74], Huang et al. used the two-dimensional (2D) deep CNN for the ECG arrhythmia classification method. By using a short-time Fourier transform, the time-domain ECG signal is transformed into an image of time–frequency spectrograms, and the outcome is applied as an input to two-dimensional CNN and achieved an accuracy of 99.0%. It has been observed that with a learning rate of 0.001 and a batch size parameter is 2500, accuracy is reached to a higher level and correspondingly decreases the average loss. In [43], the work shown is based on spiking neural networks with STDP, and RSTDP is used to train the model weights according to the spike signal timings. For real-time operation, the proposed work is appropriate. The simulation results in energy consumption of 1.78 μJ per beat in real-time classification which is 2 to 9 orders of magnitude smaller than another classification method. The method obtained an overall accuracy of 97.9%. In [44]. Wu et al. aimed at multiclass heartbeat recognition in wearable devices where BLSTM and CNN are used to deliver high accuracy with minimal network scale. The weight quantization and parameter precision reduction are adopted as network compression techniques to reduce the storage of parameters for better classification results. The method has shown the accuracy of 98.6% and 98.3% in VEB and SVEB, respectively.

In [45], a novel approach has been introduced that comprises two components deep intermediate representation (DIR) and ISVS for representation and classification respectively. MLPNN has been studied that reflects the accuracy of 99.5%. MIT-BIH Arrhythmia database and PTB diagnostic ECG database have been used. In [70], Feng et al. presented the PPNN for the multi-channel time-varying signal classification problems. Softmax classifier implemented the probabilistic classification of time-varying signals with high efficiency. Specific learning algorithms have been developed that synthesize dynamic time warping, C-means clustering, and Bayesian probability algorithm. Chinese cardiovascular disease database (CCDD) has been used which shows the accuracy and specificity of 84.92% and 84.21% respectively. In [73], 5-level wavelet decomposition was performed by using the stop band energy (SBE) optimal orthogonal wavelet filter. Here, for the classification of heartbeats, noisy data and clean data have been used. Fuzzy entropy, Renyi entropy, and fractal dimension features were extracted for accurate classification. The features are then applied to the K-nearest neighbor (KNN) classifier that achieved a maximum accuracy, sensitivity, and specificity of noisy data is 98%, 85.33%, and 98.22, respectively and for clean data, it is 98.1%, 85.63%, and 98.27%, respectively. The positive predictive value for noisy and clean data is 94.81% and 94.85%, respectively. To assist the clinicians to aid in their diagnosis, this developed system can be used in the intensive care units. In [54], Hua et al. introduced a new QRS detection algorithm followed by a deep Boltzmann machine (DBM) based classification. A template-based algorithm was used for QRS detection. MIT-BIH database has been used to verify the proposed method. When the compression ratio (CR) is 40%, it reflects the accuracy of 90% and 81.88% with MIT-BIH and recorded database respectively, and specificity of 95.52% and 94.44% respectively. In [46], the Spark Scala tool is used to simplify the machine learning algorithms. It provides the simplest ways to implement classification algorithms like decision trees, random forests, gradient boosted trees (GDB), etc. The proposed method is validated on MIT-BIH Arrhythmia and MIT-BIH Supraventricular database. For binary classification, the proposed approach achieved an accuracy of 96.75% and 97.98% using

GDB trees and the random forest algorithm, respectively. By using random forest, 98.03% accuracy is achieved for multi-class classification.

In [48], Shaker et al. introduced the data augmentation technique by using a generative adversarial network (GAN is an unsupervised method, it consists of two neural networks) to maintain the balance of the dataset. An end-to-end approach (directly classifies the heartbeat) and a two-stage hierarchical approach have been used which is based upon deep CNN. The proposed method achieved an overall accuracy, sensitivity, and specificity of 98.30%, 99.77%, and 99.23% respectively. In [49], Zu and Liu used the CNN-based ECG classification method. It has been explored by the computer-aided method. Denoising has been done by wavelet filters and worked upon the MIT-BIH arrhythmia database. Overall accuracy and specificity of 99.43% and 99.9% are achieved, respectively. The sensitivity and positive predictive values of VEB and SVEB are 99.2%, 97.5%, and 99.4%, 99.1%, respectively. To monitor long-term ECG data, the proposed system can be directly implemented on wearable devices. In [75], Sun et al. presented a novel ensemble multi-label classification process that involves several classification methods to enhance the performance of the classifier. Based on mutual information, the weights of each classifier are computed. A genetic algorithm is used to calculate the optimum threshold value. CCDD database is used in the research work with feature extraction of morphological features, time and frequency domain features, and obtained an overall accuracy of 75.2%. AKMC algorithm for feature extraction with KNN classifier and visual morphological pattern features is used which shows an overall accuracy of 97.70% by Yang and Wei [76].

Wang [60] presented the one-class support vector machine (OCSVM) with an accuracy of 98%. Rondhane et al. [65] implemented the novel loss function called focal loss with the CNN approach that enhance the classification accuracy to 98.41%, F1-score to 98.38%, and precision to 98.37%. In [59], Segmentation and extreme machine learning (ELM) based on differential evolution algorithm (DEA-ELM) with DWT approach show an accuracy of 98.4% and 97.5% respectively. Filtered and non-filtered data has been separated by CAD system through two-dimensional CNN with an accuracy of 99.23% and 99.18% respectively [58]. Song [50] presented the residual concatenate network through information-theoretic metric learning with an accuracy, sensitivity, and specificity of 88.20%, 80.45%, and 94.81% respectively. Dokur [51] revealed the DivNet with the convolutional neural network having an accuracy of 99.45%. Huang [52] shows the wavelet packet decomposition for preprocessing with FCResNet and an accuracy of 98.79%. Yan et al. [53] presented the neural network-based approach that consists of CNN and SNN attained the accuracy of 92% and 91% respectively. Lu et al. [77] experimented with 1D-CNN on the Xilinx ZC706 platform and the z-score process attained accuracy, sensitivity, specificity, and ppv of 99.10%, 99.13%, 98.59%, and 99.10% respectively. In [78], introduced a multi-interval symmetrized dot pattern method worked on image-shaped features with a weight-based relief method, and a genetic algorithm showed an accuracy of 99.74%. Kung et al. [72] proposed the wave detection algorithm with the specificity of 99.61% and 99.83% in SVEB and VEB, respectively. Yang and Wei [79] used the discrete wavelet transform approach with an ensemble classifier (KSMAX: KNN, SVM, MLP, AdaBoost, XGBoost) having an overall accuracy of 98.68%. Ganguly et al. [80] presented a unidirectionally processed multifractal detrended fluctuation analysis (u-MDFA) with BLSTM as a classifier which attained an accuracy of 97.3% and specificity in the case of SVEB and VEB is 99.2% and 98.6% respectively. In [81], introduced a bagging model with meta classifier (fusion classifier) on local and classical features having an accuracy of 95.81%. Mahmud et al. [82] presented a wavelet decomposition technique with Deep-ArrNet, PTP (Pointwise-temporal-pointwise) approach on temporal features which attained the accuracy and sensitivity of 99.28% and 99.13%, respectively. Nurmani et al. [83] introduced the Stacked Convolutional Bi-LSTM model with Bi-LSTM as a classifier and showed a specificity of 99.90%. In [84], Houssein et al. presented a marine predators algorithm (MPA) with MPA-CNN as a classifier having features like RR interval, 1D local binary pattern (LBP), and Higher-order statistics (HOS), Hermite basis function (HBF). The model attained an overall accuracy of 99.76%. A genetic algorithm having a specificity of 98.90% worked with KNN, SVM, and MLP as classifiers presented by Hammad et al. in [85]. In [86], Wang et al. proposed the tri-branch CNN model with a mixup asymmetric tri-training (MIAT) method on morphological features that classify the ECG signal in different beats. In [87], Ahmad et al. presented the multimodal image fusion (MIF) and multimodal feature fusion (MFF) models with SVM, and softmax as classifiers having a precise value of MIF and MFF is 92% and 98%, respectively. Some recently used efficient techniques for preprocessing and peak detection are summarised in [88–97].

A detailed summary of ECG arrhythmia detection with different classification approaches is presented in Table 7.

5 Challenges and Limitations

Cardiovascular disease (CVD) is one of the main causes of death all over the world [60]. Various factors like diabetes, blood pressure, hypertension, and others degrade the functioning of the heart causing severe conditions resulting in death. Due to insufficient monitoring tools, the death rate due to cardiovascular disease increases very rapidly. Early

detection and monitoring of the abnormality in the heart activity are helpful to prevent CVD. With regular checkups & treatment of heart-related problems, several diseases can be prevented by ECG diagnosis. There are various interferences in the acquisition process of ECG signals that may lead to the bad quality of the collected ECG signals [98].

Based on the literature review, ECG classification is an efficient method to classify the signal according to the heart rate and calculate the abnormality which further detects the arrhythmia. One of the key challenges is to reduce the effect of noise from the ECG signal. The accuracy of the system suffers due to noise effects which directly increase the computational complexity and correspondingly reduce the performance. Analog filters are not efficient to filter out the unwanted noises from ECG signals. Due to some limitations, digital filters are more flexible than analog filters for denoising the ECG signal [99]. It has been observed that various researchers implemented different filters like low pass filter (LPF), high pass filter (HPF), band pass filter (BPF), and notch filter for noise removal. LPF [100] performs better for denoising as it removes all high-frequency noise like PLI, and EMG but also affects the QRS complex and attenuates the ECG signal characteristics. To overcome this, researchers prefer HPF and BPF to eliminate noises like EMG, PLI, and Baseline wander. To remove noise from the images and signals, the median filter can also be used [8]. In literature, time-domain approaches like a synchronized averaging filter, moving averaging filter, and frequency domain approaches, adaptive filtering has been used to denoise the signal. Various algorithms are implemented to design the adaptive filter like least mean square (LMS), recursive least square (RLS), and normalized least mean square (NLMS). The overall combined time–frequency domain approach showed better results for denoising the signal. Various techniques have to be implemented to eliminate the noise and artifacts. Another challenge is the detection of the QRS complex as it shows cardiac dysfunction due to multiple premature QRS complexes which reflects the complexity of the detection process [8]. Several thresholding techniques can be used to detect QRS complex but in all techniques, the adaptive threshold technique performs better which increases the probability of detecting QRS complexes. Artificial neural networks are also useful and helpful in detecting QRS complex. One more important factor is the selection of sampling frequency. An appropriate selection of sampling frequency improves the accuracy of the detection algorithm of the QRS complex [101]. If the sampling frequency is low, produces an error due to the superimposition of high-frequency components on the low-frequency components. Another key challenge is to develop a real-time, unobtrusive method to monitor heart conditions with physical activities. Heart rate can be calculated through wearable devices like bands or watches but the accuracy of wearable devices can be suffered through motion artifacts [102]. To address this problem, numerous solutions have been proposed with advanced signal processing algorithms like adaptive noise cancellation methods, but it is still an obstruction to overcome with this artifact. Unobtrusive monitoring devices can be implemented in different ways, by inserting the wearable devices or sensors in different accessories, or by fixing the sensors in household objects like chairs, tables or furniture, etc. Sometimes, a problem of blurred ECG image occurs which shows the misdiagnosis while detecting the arrhythmia. To overcome this, high spatial resolution imaging methods have to be developed or high investment in imaging machines will help out. From previous papers, it has been seen that the 1D-ECG signal with CNN reflects high accuracy but at a cost of computational complexity. So, to enhance the system performance, the 1D-ECG signal is converted into a 2D-ECG image and then fed to the classifiers in the convolutional neural network which shows the maximum accuracy with the least error and computational complexity. According to researchers, different algorithms with classifiers have been used to enhance the performance of the system. Several models like SNN based classifier [43], KNN, SVM classifier [73, 76], Gradient Boosted Tree and Random Forest [46], ensemble SVM [98], MOWPT [52], FCResNet model [52], General Sparse Neural Network (GSNN) [77], DeepArrNet (1D-CNN) [78] have been used which reflects the high accuracy but as compared to all models, 2D-ECG image with CNN [24, 48, 51, 58, 74] shows the highest accuracy with least computational complexity. This technique performs best in all other above techniques that represent the maximum accuracy with the least computational complexity.

6 Conclusions

The review shows that AI-based computational techniques have been widely developed to analyze the ECG signal and detect cardiovascular disease. For the prevention of cardiovascular diseases, early detection and treatment are necessary in today's world. The emerging technology in health informatics which deals with data acquisition, processing, and detection can come with new opportunities to prevent from CVDs. AI-based techniques provide accurate outcomes in the classification of ECG signals to detect arrhythmia. Various steps are involved in the process of classification. Denoising is the crucial and very important step in ECG analysis for which several techniques have to be implemented. With the identification of morphological, and temporal features from a denoised ECG signal, the CVDs can be detected effectively. To verify the characteristics of ECG signals, QRS detection techniques have been discussed. Different thresholding techniques have been implemented for detection purposes. This paper summarizes the techniques

Table 7 Summary of ECG arrhythmia detection with different classification approaches available in the literature

S.NO	Authors	Classifier	Dataset	Material/ methodology/ process used	Algorithm	Features	Accuracy (%)	Sensitivity (%)	Specificity (%)	Positive predictive value (ppv)(%)	Precision/Recall/F score (%)
1	Zaunseder et al. [1]	LDA or KNN, MLP	MIT-BIH arrhythmia database	Wavelet transform, PCA	Sequential forward floating search (SFFS)	Temporal (heart rate feature), Morphological (amplitude, peak width), Statistical (variances)	LDA = 84.6 MLP = 89	LDA = 76.5 MLP = 80.2	NR	SVEB = 38.5 VEB = 81.9	NR
2	Barni et al. [2]	NN, Linear branching program (LBP); Decision tree, QDF classifier	Physiobank database	Hybrid approach (Homomorphic encryption, garbled circuits); Secure signal processing (SSP classifier)	Levenberg–Marquardt regularized learning algorithm	Smaller feature vector (4 instead of 15 components of feature vector)	86.3	NR	NR	NR	NR
3	Alvarado and lakshminarayan [3]	LDA	MIT-BIH arrhythmia database	Time-based Integrate and fire (IF) Sampler; Time-based encoding	Use a simple binning approach (integer addition)	NR	NR	SVEB = 86.19 VEB = 92.43	SVEB = 97.45 VEB = 99.6	SVEB = 56.68 VEB = 94.82	NR
4	Moran et al. [4]	Logistic Regression	MIT-BIH Database	Reservoir Computing, Delay System	Ring configuration	NR	98.43	84.83	97.75	88.75	NR
5	Banerjee and Mitra [5]	Threshold-based classification over leads II, III, aVF	PTBDB ECG diagnostic database	Discrete wavelet transform-based decomposition	Cross wavelet transform	NR	97.6	97.3	98.8	NR	NR
6	Lee et al. [6]	Biosignal processor (CBCTA and HPSDM)	MIT-BIH database; Technology TSMC 0.18 µm standard CMOS process	Wavelet transform	HPSDM, low power super-regenerative on–off keying transceiver, DSP	Power consumption of body end circuit (BSP and transmitted circuit) = 586.5µw CBCTA- 17 µw HPSDM- 37.5 µw	Beat detection = 99.44 ECG classification = 97.25	NR	NR	NR	NR
7	Huang and Zhang [7]	SVM	The local hospital, courtesy of the SiWei medical company (Shanghai, China) & SiWei Remote ECG diagnostic center	Tensor learning approach; Short-time fourier transform	Generalized tensor rank one discriminant analysis (GTR1DA)	NR	N- 99.86 L- 99.73 R- 100 V- 100 S- 100 E- 100	NR	NR	NR	NR

Table 7 (continued)

S.NO	Authors	Classifier	Dataset	Material/ methodology/ process used	Algorithm	Features	Accuracy (%)	Sensitivity (%)	Specificity (%)	Positive predictive value (ppv)(%)	Precision/Recall/F score (%)
8	Rahman et al. [10]	RF, SVM	Hypertrophic Cardiomyopathy (HCM Patients), ischemic cardiomyopathy (ICD Patients)	fivefold cross-validation	NA	Temporal and Morphological features	NR	NR	0.93	NR	Precision = 0.84, recall = 0.89, F score = 0.86
9	Shadmand and Mashoufi [13]	Block-based Neural Network (BBNN)	MIT-BIH arrhythmia database	ANN	PSO algorithm	Morphological and Temporal features	VEB = 98 SVEB = 97.4	VEB = 87.4 SVEB = 58.6	VEB = 98.8 SVEB = 99	VEB = 88.9 SVEB = 71.3	NR
10	Oster et al. [15]	SVM classifier	MIT-BIH Arrhythmia and Incart databases	Switching kalman filter approach (SKF); X factor	Denoising by non-linear Bayesian filtering	NR	NR	V' classification: X-factor = 98.6, V classification: X-factor = 99.1	V' classification X-factor = 99.9 V classification: X-factor = 99.96	NR	MIT-BIH: F1 = 98.3 INCART: F1 = 99.5
9	Gutta and Cheng [56]	Probabilistic non-linear kernel classifier	MIT-BIH NSR database from Physionet	Multitask learning approach; Data preprocessing	Fast optimization algorithm	Non-fiducial feature	NA	NA	NA	NA	NA
11	Padmavathi and Ramakrishna [57]	KSVM and KNN	MIT-BIH Atrial Fibrillation Database	Burg's method, Atrial Fibrillation (AF)	Autoregressive (AR) coefficients	Morphological features	Burg + KNN = 100 Burg + SVM = 100	NR	NR	NR	NR
12	Kiranyaz et al. [36]	CNN	MIT-BIH arrhythmia database, Patient specific data	PCA	1D CNN Approach with back propagation epochs	NR	VEB = 99 SVEB = 97.6	VEB = 93.9 SVEB = 63.5	VEB = 98.9 SVEB = 99.2	VEB = 90.6 SVEB = 63.5	NR
13	Park et al. [37]	Cascaded RF classifier	MIT-BIH Arrhythmia Database	Adaptive feature extraction	KNN algorithm	Heartbeat morphology features	97.34	91.23	99.81	97.89	NR
14	Teijeiro et al. [20]	Rule-based classifier	MIT-BIH Arrhythmia Database	Abductive interpretation of the signal	QRS clustering algorithm	Morphological and rhythm features	NR	VEB = 92.82 SVEB = 85.10	NR	VEB = 92.23 SVEB = 84.5	NR
15	Eftestol et al. [14]	LDA, QDA, ANN classifier with Bayesian regularization	1631 3-s ECG segments with clinical rhythm annotations, from 298 out-of-hospital cardiac arrest patients	Gaussian mixture model (GMM), Resuscitation; Nested cross-validation architecture for feature selection & model assessment	47 wavelet- and time domain-based features; wrapper-based feature selection architecture	Discrete Wavelet Transform = 32 features, Classical features = 12, R wave detection = 3 features	78.5	AS = 88.7 PEA = 68.9 PR = 65.9 VF = 86.2 VT = 78.8	NR	AS = 91 PEA = 70.4 PR = 69 VF = 83.8 VT = 72.9	NR

Table 7 (continued)

S.NO	Authors	Classifier	Dataset	Material/ methodology/ process used	Algorithm	Features	Accuracy (%)	Sensitivity (%)	Specificity (%)	Positive predictive value (ppv)(%)	Precision/Recall/F score (%)
17	Satija et al. [17]	NR	MIT-BIHA database, MIT-BIHSTC, PTBDECG dB, Fantasia database, PCICC2011 database	CEEMD	CEEMD	Temporal features (amplitude, zero crossings etc.)	Detection of noise = 98.9 Classification = 97.38	Detection of noise = 99.12 Classification = 98.93	Detection of noise = 98.56	Classification = 98.39	NR
18	Patro and Kumar [18]	ANN for pattern recognition classifier	MIT-BIH ECG ID database signals	Time–frequency approach and Wavelet transform approach	Denoising by cascaded filter	Morphological and Temporal features	98.14	NR	98.9	NR	Precision = 93.75 Recall = 88.88 F Score = 88.53
19	Li et al. [19]	CNN classifier	MIT-BIH arrhythmia database	Relu activation function	Generic CNN(GCNN), TDCNN	Hierarchical features	Overall = 96.89 VEB = 98.8 SVEB = 98.3	VEB = 96.3 SVEB = 72.9	VEB = 99 SVEB = 99.8	VEB = 92 SVEB = 95.6	NR
20	Sannino and Pietro [38]	DNN Classifier	MIT-BIH arrhythmia database	Deep learning approach	ANN	Temporal features	99.68	99.48	99.83	NR	NR
21	Ayar and Sabamoniri [21]	Decision tree	UCI arrhythmia dataset	Hybrid model	Decision tree with C4.5, Genetic algorithm	Optimal features	2 class = 86.96 16 class = 78.76	2 class = 88.88 16 class = 76.36	2 class = 86.55 16 class = 78.69	NR	Precision 2class = 88.8 16class = 79.2 Recall 2 class = 86.5 16 class = 76.3 F1 score 2class = 88.8 16class = 77.7
22	Sayantan et al. [22]	Linear SVM	MIT-BIH Arrhythmia Database and MIT-BIH SVDB	Gaussian-Bernoulli deep belief network	NR	NR	MIT-BIH SVEB- 99.5% VEB- 99.4 SVDB Database SVEB- 97.5 VEB- 98.6	MIT-BIH SVEB = 93.7 VEB = 98.3 SVDB Database SVEB- 81.6 VEB- 87.5	MIT-BIH SVEB = 99.6 VEB = 99.42 SVDB Database SVEB = 98.6 VEB = 99.2	NR	NR
23	Celin and Vasanth [23]	SVM, Adaboost, ANN and Naïve Bayes classifier	MIT-BIH database, ECG –ID database	Preprocessed by using filtering method; Butter worth filter	Peak detection algorithm	Statistical parameters (mean, standard deviation, root mean square, pulse transit time)	SVM = 87.5 Adaboost = 93 ANN = 94 Naïve bayes = 99.7	SVM = 75	SVM = 100	NR	NR
24	Zhai and Tin [24]	2D CNN classifier	MIT-BIH arrhythmia database	Dual beat ECG coupling matrix; Bandpass filter	NR	Morphological and dynamic features	VEB = 98.6 SVEB = 97.5	VEB = 93.8 SVEB = 76.8	VEB = 99.2 SVEB = 98.7	VEB = 92.4 SVEB = 74	NR

Table 7 (continued)

S.NO	Authors	Classifier	Dataset	Material/ methodology/ process used	Algorithm	Features	Accuracy (%)	Sensitivity (%)	Specificity (%)	Positive predictive value (ppv)(%)	Precision/Recall/F score (%)
25	Iqbal et al. [55]	NR	MIT-BIH arrhythmia database	Autoregressive modeling and Hidden Markov model	Model-driven architecture, MDA	Morphological features	NR	NR	NR	NR	NR
26	Zhu et al. [25]	SVM-RBF	MIT-BIH arrhythmia database	PCA and DTW	fivefold cross-validation	Morphological and Segment features	97.80	N = 99.27 S = 87.47 V = 94.71 F = 73.88	NR	N = 98.48 S = 95.25 V = 95.22 F = 86.09	NR
27	Li and Li [39]	Local deep field (LDF) (local regionalization & deep learning)	MIT-BIH arrhythmia database	L2 norm penalty regularization; tenfold cross-validation	MLP	NR	99.01	NR	NR	NR	NR
28	Shensheng et al. [26]	DNN; Patient-independent 1D classifier	MIT-BIH arrhythmia database	K fold cross-validation	NR	Raw ECG	99.70	VEB = 93.7 SVEB = 77.3 PVC = 97.68	VEB = 98.8 SVEB = 97.7 PVC = 99.89	NR	NR
29	Niu et al. [27]	MPCNN	MIT-BIH arrhythmia database	SBCX approach	ADAM optimizer, SAX algorithm	Morphological and rhythm features	96.4	N = 98.9 S = 76.5 V = 85.7	NR	N = 97.4 S = 76.6 V = 94.1	F1 Score N = 98.1 S = 76.6 V = 89.7
30	Lynn et al. [28]	1D-CNN, Deep RNN	ECG-ID database and MIT-BIH Arrhythmia Database (MITDB)	BGRU model	Bidirectional RNN with LSTM cell (BLSTM) and GRU cell (BGRU)	NR	98.5 (ECGID = 98.6 MITDB = 98.4)	NR	NR	NR	NR
31	Saadatnejaad et al. [40]	Multiple LSTM RNN	MIT-BIH ECG arrhythmia database	GRU	Wavelet transform and multiple LSTM RNN	Classical features, Wavelet features	VEB = 99.2 SVEB = 98.3	VEB = 93 SVEB = 66.9	VEB = 99.8 SVEB = 99.8	VEB = 98.2 SVEB = 95.7	VEB, F1 = 95.5 SVEB, F1 = 78.8
32	Wang et al. [41]	SDAE + Bi-LSTM classifier	MIT-BIH Arrhythmias Database, MIT-BIH SVDB, and MIT-BIH NSTDB	SDAE	NA	Temporal features	98.85	96.32	99.86	NR	NR
33	Rahhal et al. [42]	Weak linear classifier (WLC), SVM classifier	MIT-BIHA database, INCART, MIT-BIH SVDB	Focal loss (reshaping cross-entropy loss); Generative and discriminative (DenseNet block) module	Dense convolutional networks (DCN)	NR	VEB = 97 SVEB = 74	VEB = 85 SVEB = 67	VEB = 98 SVEB = 98	VEB = 79 SVEB = 51	NR

Table 7 (continued)

S.NO	Authors	Classifier	Dataset	Material/ methodology/ process used	Algorithm	Features	Accuracy (%)	Sensitivity (%)	Specificity (%)	Positive predictive value (ppv)(%)	Precision/Recall/F score (%)
34	Huang et al. [74]	2D-CNN	MIT-BIH arrhythmia database	STFT	2D CNN, STFT-based spectrogram	NR	99 Average loss = 4.1	NR	NR	NR	NR
35	Amirshahi and Hashemi [43]	SNN	MIT-BIH arrhythmia database	STDP, and RSTDP	Spiking neural networks (SNN) based algorithm	Temporal features	97.9	80.2	99.8	97.3	F1 score = 88
36	Wu et al. [44]	CNN, BLSTM	MIT-BIH arrhythmia database	BLSTM	SMIC 40LL CMOS process, Energy efficiency-3.52 GOPS/mw	Morphological features	VEB = 98.6 SVEB = 98.3	VEB = 91.8 SVEB = 80.6	VEB = 99.4 SVEB = 99.1	VEB = 94 SVEB = 82.4	F1: VEB = 92.9 SVEB = 81.5
37	Li [45]	ISVS for classification	MIT-BIH arrhythmia database, PTB Diagnostic ECG Database	Deep Intermediate Representation (DIR)	MLPNN algorithm	NR	99.5	NR	NR	NR	NR
38	Feng et al. [70]	Softmax classifier	CCDD database	PPNN model	DTW, C means clustering & Bayesian probability algorithm	Morphological features	84.92	NR	84.21	NR	Recall = 85.13 F1 score = 86.03
39	Sharma et al. [73]	KNN classifier	MIT-BIH ECG database	Wavelet filter bank (5-level wavelet decomposition)	Optimal orthogonal wavelet filter	Fuzzy entropy, Renyi entrop,y and fractal dimension features	Noisy data = 98 Clean data = 98.1	Noisy data = 85.33 Clean data = 85.63	Noisy data = 98.22 Clean data = 98.27	Noisy data = 94.81 Clean data = 94.85	Noisy data = 89.45 Clean data = 89.65
40	Hua et al. [54]	Deep Boltzmann machine (DBM) classifier	MIT-BIH database, Our database	Compressive sensing technique	Template-based algorithm	NR	When CR = 40 MIT-BIH = 90 Our = 81.88	When CR = 40 MIT-BIH = 86.02 Our = 75.47	When CR = 40 MIT-BIH = 95.52 Our = 94.44	NR	NR
41	Alarsan and Younes [46]	Binary and Multi Classification (GDB and RF)	MIT-BIH Arrhythmia and MIT-BIH SVDB	ML-libs and Scala language on Apache Spark framework	Decision Tree, Random Forests, Gradient-Boosted Trees (GDB)	Temporal and Morphological features	For binary classification, GDB = 96.75 RF = 97.98 For multiclass classification, RF = 98.03	NR	NR	NR	NR
42	Shaker et al. [48]	Deep CNNs	MIT-BIH arrhythmia database	Data augmentation technique using GANs	An end-to-end approach and a two-stage hierarchical approach	NR	98.30	99.77	99.23	NR	Precision = 90
43	Xu, Liu [49]	CNN classifier	MIT-BIH arrhythmia database	Denoising by wavelet filters	Computer aid method	NR	99.43	VEB = 99.2 SVEB = 97.5	99.9	VEB = 99.4 SVEB = 99.1	NR
44	Sun et al. [75]	Ensemble classifier	CCDD database	Multi-label ECG classification	Genetic algorithm	Morphological, time domain, and frequency domain features	75.2	NR	NR	NR	F1 = 75.2

Table 7 (continued)

S.NO	Authors	Classifier	Dataset	Material/ methodology/ process used	Algorithm	Features	Accuracy (%)	Sensitivity (%)	Specificity (%)	Positive predictive value (ppv)(%)	Precision/Recall/F score (%)
45	Yang and Wei [76]	KNN classifier	MIT-BIT arrhythmia database	Clustering-based feature extraction	AKMC algorithm for extracting this feature	Parameter + Visual morphological pattern (VMP-QRS) feature	Overall = 97.70 Normal = 97.79 Left = 99.50 Right = 99.59 V = 97.69 Artery A = 89.7 Paced P = 99.92	NR	NR	NR	NR
46	Wang et al. [60]	Tree classifier using OCSVM	PCICC2011 database	Automatic classification of ECG data quality by channel	One class support vector machine (OCSVM)	Electrode shedding C3, Serious noise interference C2, Partial noise interference C1, High-quality signal C0	Overall = 98 Detection accuracy of C3-93.22, C2-90, C1-89.22, C0-97.19	97.9	98.9	NR	NR
47	Pandey et al. [98]	Ensemble SVM	MIT-BIH arrhythmia database	Ensemble	NA	Wavelets, high-order statistics, R-R intervals and morphological features	94.40	65.26	93.25	NR	Precision = 69.11 F score = 66.24
48	Romdhane et al. [65]	CNN model	MIT-BIH and INCART datasets	CNN + Loss function (Focal loss)	Segmentation algorithm	Morphological features	98.41	NR	NR	NR	Precision = 98.37 Recall = 98.41 F1 score = 98.38
49	Diker et al. [59]	DEA-ELM	PTB Diagnostic ECG database	ELM, Pan-Tompkins technique (PTT), and DWT approach	DEA	Morphological and statistical features	97.5	NR	97.7	NR	NR
50	Huang and Wu [58]	CAD system for classification through 2D-CNN	MIT-BIH AF Dataset and MIT-BIH NSR Dataset	CNN; two-dimensional ECG image	Computer-aided diagnosis system (CAD)	NR	Filtered = 99.23 Non filtered = 99.18	Filtered = 99.71 Non filtered = 99.31	Filtered = 98.66 Non filtered = 99.03	NR	NR
51	Song et al. [50]	Residual + concatenate	MIT-BIH arrhythmia database	Residual-concatenate network	Information-theoretic metric learning	Low-level and high-level features of short-term ECG	88.20	80.45	94.81	82.48	NR
52	Dokur and Olmez [51]	CNN, DNN	MIT-BIH arrhythmia database	Walsh functions; DivNet	1D signal into 2D images	NR	1D ECG signal = 99.45 ECG image = 98.7	NR	NR	NR	NR
53	Huang et al. [52]	CNN, (FCResNet)	MIT-BIH arrhythmia database	Wavelet packets decomposition (WPD) for preprocessing	Maximal overlap wavelet packet transform (MOWPT)	NR	98.79	NR	NR	NR	Precision = 99.39 Recall = 95.16 F1 score = 97.23
54	Yan et al. [53]	CNN and SNN	MIT-BIH Arrhythmia Database	Neural network-based approach	NR	Morphological features	CNN = 92 SNN = 91	NR	NR	NR	The average power of SNN = 0.077 W The average power of CNN = 0.67 W

Table 7 (continued)

S.NO	Authors	Classifier	Dataset	Material/ methodology/ process used	Algorithm	Features	Accuracy (%)	Sensitivity (%)	Specificity (%)	Positive predictive value (ppv)(%)	Precision/Recall/F score (%)
55	Lu et al. [77]	1-D CNN	MIT-BIH arrhythmia database	Xilinx ZC706 platform	Z-score process	NR	99.10	99.13	98.59	99.10	NR
56	Yang and Yan [78]	SVM	MIT-BIH arrhythmia database	Multi-interval symmetrized dot pattern method (MSDP)	Weight-based ReliefF method, genetic algorithm	Image shaping features	99.74	99.42	99.83	NR	NR
57	Kung et al. [72]	Random forest	MIT-BIH arrhythmia database, QT database, CinC 2017 A-fib dataset	NR	Wave detection algorithm	NR	SVEB = 98.67 VEB = 99.63	SVEB = 75.38 VEB = 96.71	SVEB = 99.61 VEB = 99.83	SVEB = 88.71 VEB = 97.43	F1 score: SVEB = 81.05 VEB = 97.07
58	Yang and Wei [79]	Ensemble classifier (KSMAX: KNN, SVM, MLP, AdaBoost, XGBoost)	MIT-BIH arrhythmia database	NR	Discrete wavelet transform	Morphological features	Overall = 98.68 N = 98.75 RBBB = 99.77 LBBB = 99.70 APC = 98.57 PVC = 99.94	NR	NR	NR	NR
59	Ganguly et al. [80]	Bi- LSTM	MIT-BIH arrhythmia database	u-MDFA	Discrete wavelet transform	NR	97.3	SVEB = 96.5 VEB = 97	SVEB = 99.2 VEB = 98.6	NR	NR
60	Essa and Xie [81]	Meta classifier (Fusion classifier)	MIT-BIH arrhythmia database	Bagging model (CNN-LSTM, RR-HOS-LSTM)	NR	Local features, classical features	95.81	N = 98.03 SVEB = 65.51 VEB = 93.91 F = 19.33	NR	N = 97.48 SVEB = 66.19 VEB = 94.5 F = 41.67	NR
61	Mahmud et al. [82]	1D CNN	MIT-BIH arrhythmia database, PTB dB	DeepArrNet, PTP (Pointwise-temporal-pointwise)	Wavelet decomposition technique	Temporal features	99.28	99.13	NR	99.08	F1 score = 99.11
62	Nurmaini et al. [83]	Bi-LSTM classifier	QT database	Stacked Convolutional Bi-LSTM model	NR	NR	99.83	98.82	99.90	NR	Precision = 98.86 F1 score = 98.84
63	Houssein et al. [84]	MPA-CNN classifier	MIT-BIH arrhythmia, EUROPEAN ST-T dataset (EDB), INCART database	NR	Marine predators algorithm (MPA)	RR interval, 1D LBP, HOS, HBF	Overall = 99.76 MIT-BIH = 99.31 EDB = 99.76 INCART = 99.4	Overall = 99.49	NR	NR	MIT-BIH = 99.31 EDB = 99.76 INCART = 99.4
64	Hammad et al. [85]	KNN, SVM, MLP classifier	MIT-BIH arrhythmia	Deep learning model	Genetic algorithm (GA)	NR	98.0	99.70	98.90	95.80	F1 score = 89.7

Table 7 (continued)

S.NO	Authors	Classifier	Dataset	Material/ methodology/ process used	Algorithm	Features	Accuracy (%)	Sensitivity (%)	Specificity (%)	Positive predictive value (ppv)(%)	Precision/Recall/F score (%)
65	Wang et al. [86]	CNN	MITDB, SVDB	MIAT method	Tri branch CNN model	Morphological features	NR	N = 99.4 S = 77.2 V = 90.6 F = 66.8	NR	N = 98.3 S = 93.4 V = 94.4 F = 75.5	F1 score: N = 98.9 S = 84.5 V = 92.4 F = 70.9
66	Ahmad et al. [87]	SVM and softmax classifier	MIT-BIH arrhythmia, PTB database	MIF and MFF	NR	Gramian Angular Field (GAF), Recurrence Plot (RP), and Markov Transition Field (MTF)	MIT-BIH Arrhythmia = 99.7 MI classification = 99.2	NR	NR	NR	Precision: MIF = 92 MFF = 98 Recall: MIF = 94 MFF = 98

presented in the literature for arrhythmia detection and classification purposes with regular measurement of heart rate. The performance of different methods on several databases is listed in tabular form. A Convolutional neural network reflects better accuracy in arrhythmia detection. Even though, our healthcare system requires wearable devices to be implemented for continuous monitoring of cardiac health. Some limitations have been observed in this area by the researchers. Better equipment is necessary for excellent performance in the analysis of ECG signals. Despite many challenges, these computational methods are becoming a powerful tool for medical advances.

Declarations

Conflict of interest All Authors of this work declare no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Mar T, Zaunseder S, Martinez JP, Llamedo M, Poll R (2011) Optimization of ECG classification by means of feature selection. IEEE Trans Biomed Eng 58(8):2168–2177
2. Barni M, Failla P, Lazzeretti R, Sadeghi AR, Schneider T (2011) Privacy-preserving ECG classification with branching programs and neural networks. IEEE Trans Inf Forensics Secur 6(2):452–468
3. Alvarado AS, Lakshminarayan C, Principe JC (2012) Time-based compression and classification of heartbeats. IEEE Trans Biomed Eng 59(6):1641–1648
4. Moran ME, Soriano MC, Fischer I, Mirasso CR (2012) Electrocardiogram classification using reservoir computing with logistic regression. IEEE J Biomed Health Inform 19(3):892–898
5. Banerjee S, Mitra M (2013) Application of cross wavelet transform for ECG pattern analysis and classification. IEEE Trans Instrum Meas 63(2):326–333
6. Lee SY, Hong JH, Hsieh CH, Liang MC, Chien SYC, Lin KH (2014) Low-power wireless ECG acquisition and classification system for body sensor networks. IEEE J Biomed Health Inform 19(1):236–246
7. Huang K, Zhang L (2014) Cardiology knowledge free ECG feature extraction using generalized tensor rank one discriminant analysis. EURASIP J Adv Signal Process. https://doi.org/10.1186/1687-6180-2014-2
8. Tripathi PM, Kumar A, Komaragiri R, Kumar M (2022) A review on computational methods for denoising and detecting ECG signals to detect cardiovascular diseases. Arch Comput Methods Eng 29:1875–1914
9. Wasimuddin M, Elleithy K, Abuzneid A, Faezipour M, Abuzaghleh O (2020) Stages based ECG signal analysis from traditional signal processing to machine learning approaches: a survey. IEEE Access 8:177782–177803
10. Rahman QA, Tereshchenko LG, Kongkatong M, Abraham T, Abraham MR, Shatkay H (2015) Utilizing ECG-based heartbeat

classification for hypertrophic cardiomyopathy identification. IEEE Trans NanoBiosci 14(5):505–512
11. Murawwata S, Asifd HM, Iiaza S, Malike MI, Raahemifarb K (2021) Denoising and classification of Arrhythmia using MEMD and ANN. Alexandria Eng J 61(4):2807–2823
12. Morales J, Moeyersons J, Armanac P, Orini M, Faes L, Overeem S, Gilst MV, Dijk JV, Huffel SV, Bailon R, Varon C (2021) Model based evaluation of methods for respiratory sinus arrhythmia estimation. IEEE Trans Biomed Eng 68:1882–1893
13. Shadmand S, Mashoufi B (2015) A new personalized ECG signal classification algorithm using Block-based Neural Network and Particle Swarm Optimization. Biomed Signal Process Control 25:12–23
14. Rad AB, Eftestol T, Engan K, Irusta U, Kvaloy JT, Johansen JK, Wik L, Katsaggelos AK (2017) ECG-based classification of resuscitation cardiac rhythms for retrospective data analysis. IEEE Trans Biomed Eng 64(10):2411–2418
15. Oster J, Behar J, Sayadi O, Nemati S, Johnson AEW, Gari D, Clifford GD (2015) Semi-supervised ECG ventricular beat classification with novelty detection based on switching Kalman filters. IEEE Trans Biomed Eng 62(9):2125–2134
16. Hesar HD, Mohebbi M (2020) An Adaptive Kalman Filter Bank for ECG Denoising. IEEE J Biomed Health Inform. https://doi.org/10.1109/JBHI.2020.2982935
17. Satija U, Ramkumar B, Manikandan MS (2017) Automated ECG noise detection and classification system for unsupervised healthcare monitoring. IEEE J Biomed Health Inform 22(3):722–723
18. Patro KK, Kumar PR (2017) Effective feature extraction of ECG for biometric application. Int Conf Adv Comput Commun 115:296–306
19. Li Y, Pang Y, Wang J, Li X (2018) Patient-specific ECG classification by deeper CNN from generic to dedicated. Neurocomputing. https://doi.org/10.1016/j.neucom.2018.06.068
20. Teijeiro T, Felix P, Presedo J, Castro D (2016) Heartbeat classification using abstract features from the abductive interpretation of the ECG. IEEE J Biomed Health Inform 22(2):409–420
21. Ayar M, Sabamoniri S (2018) An ECG-based feature selection and heartbeat classification model using a hybrid heuristic algorithm. IEEE Access 13:167–175. https://doi.org/10.1016/j.imu.2018.06.002
22. Sayantan G, Kien PT, Kadambari KV (2018) Classification of ECG beats using deep belief network and active learning. Med Biol Eng Comput. 56:1887–1898
23. Celin S, Vasanth K (2018) ECG signal classification using various machine learning techniques. J Med Syst. https://doi.org/10.1007/s10916-018-1083-6
24. Zhai X, Tin C (2018) Automated ECG classification using dual heartbeat coupling based on convolutional neural network. IEEE Access 6:27465–27472. https://doi.org/10.1109/ACCESS.2018.2833841
25. Zhu W, Chen X, Wang Y, Wang L (2018) Arrhythmia recognition and classification using ECG morphology and segment feature analysis. IEEE/ACM Trans Comput Biol Bioinform 16(1):131–138
26. Xu SS, Mak MW, Cheung CC (2018) Towards end-to-end ECG classification with raw signal extraction and deep neural networks. IEEE J Biomed Health Inform 23(4):1574–1584
27. Niu J, Tang Y, Sun Z, Zhang W (2019) Inter-patient ECG classification with symbolic representations and multi-perspective convolutional neural networks. IEEE J Biomed Health Inform 24(5):1321–1332
28. Lynn HM, Pan SB, Kim P (2019) A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks. IEEE Access 7:145395–145405
29. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE (2000) Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. Circulation 101(23):e215–e220. Physiobank database: https://www.physionet.org/
30. Moody GB, Mark RG (2001) The impact of the MIT-BIH Arrhythmia Database. IEEE Eng Med Biol 20(3):45–50. MIT-BIH Arrhythmia database: https://www.physionet.org/
31. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 101(23):e215–e220. MIT-BIH Normal Sinus Rhythm database. https://www.physionet.org/, https://physionet.org/content/nsrdb/1.0.0/
32. Moody GB, Mark RG (1983) A new method for detecting atrial fibrillation using R-R intervals. Comput Cardiol 10:227–230. MIT-BIH Atrial Fibrillation Database: https://www.physionet.org/
33. Bousseljot R, Kreiseler D, Schnabel A (1995) Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. Biomedizinische Technik, Band 40, Ergänzungsband 1:S 317. Physikalisch-Technische Bundesanstalt diagnostic 12-lead ECG database (ptbdb): https://www.physionet.org/
34. Iyengar N, Peng C-K, Morin R, Goldberger AL, Lipsitz LA (1996) Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. Am J Physiol 271:1078–1084
35. Lugovaya TS (2005) Biometric human identification based on electrocardiogram. [Master's thesis] Faculty of Computing Technologies and Informatics, Electrotechnical University "LETI", Saint-Petersburg, Russian Federation. MIT-BIH ECG ID database signals: https://www.physionet.org/
36. Kiranyaz S, Ince T, Gabbouj M (2016) Real-time patient-specific ECG classification by 1-D convolutional neural networks. IEEE Trans Biomed Eng 63(3):664–675
37. Park J, Kang M, Gao J, Kim Y, Kang K (2017) Cascade classification with adaptive feature extraction for arrhythmia detection. J Med Syst. https://doi.org/10.1007/s10916-016-0660-9
38. Sannino G, Pietro GD (2018) A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. IEEE Access 86:446–455
39. Li W, Li J (2018) Local deep field for electrocardiogram beat classification. IEEE Sens J 18(4):1656–1664
40. Saadatnejad S, Oveisi MH, Hashemi M (2019) LSTM-based ECG classification for continuous monitoring on personal wearable devices. IEEE J Biomed Health Inform 29(2):515–523
41. Wang EK, Zhang X, Pan L (2019) Automatic classification of CAD ECG signals with SDAE and bidirectional long short-term term network. IEEE Access 7:182873–182880
42. Rahhal MMA, Bazi Y, Almubarak H, Alajlan N, Zuair MA (2019) Dense convolutional networks with focal loss and image generation for electrocardiogram classification. IEEE Access 7:182225–182237
43. Amirshahi A, Hashemi M (2019) ECG classification algorithm based on STDP and R-STDP neural networks for real-time monitoring on ultra-low-power personal wearable devices. IEEE Trans Biomed Circuits Syst 13(6):1483–1493
44. Wu J, Li F, Chen Z, Pu Y, Zhan M (2019) A neural network based ECG classification processor with exploitation of heartbeat similarity. IEEE Access 7:172774–172782
45. Li W (2019) Deep intermediate representation and in-set voting scheme for multiple-beat electrocardiogram classification. IEEE Sens J 19(16):6895–6904
46. Alarsan FI, Younes M (2019) Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. J Big Data. https://doi.org/10.1186/s40537-019-0244-x

47. PCICC2011 database: https://www.physionet.org/, PhysioNet/Computing in Cardiology Challenge 2011. http://www.physionet.org/challenge/2011/.
48. Shaker AM, Tantawi M, Shedeed HA, Tolba MF (2020) Generalization of convolutional neural networks for ECG classification using generative adversarial networks. IEEE Access 8:35592–35605
49. Xu X, Liu H (2020) ECG heartbeat classification using convolutional neural networks. IEEE Access 8:8614–8619
50. Song X, Yang G, Wang K, Huang Y, Yuan F, Yin Y (2020) Short term ECG classification with residual-concatenate network and metric learning. Multimed Tools Appl 79:22325–22326
51. Dokur Z, Olmez T (2020) Heartbeat classification by using a convolutional neural network trained with Walsh functions. Neural Comput Appl 32:12515–12534
52. Huang JS, Chen BQ, Zeng NY, Cao XC, Li Y (2020) Accurate classification of ECG arrhythmia using MOWPT enhanced fast compression deep learning networks. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-020-02110-y
53. Yan Z, Zhou J, Wong WF (2021) Energy-efficient ECG classification with spiking neural network. Biomed Signal Process Control 63:102170. https://doi.org/10.1016/j.bspc.2020.102170
54. Hua J, Xu Y, Tang J, Liu J, Zhang J (2020) ECG heartbeat classification in compressive domain for wearable devices. J Syst Architect 104:101687
55. Iqbal U, Wah TY, Rehman MHU, Mastoi QUA (2018) Usage of model-driven environment for the classification of ECG features: a systematic review. IEEE Access 6:23120–23136
56. Gutta S, Cheng Q (2015) Joint feature extraction and classifier design for ECG based biometric recognition. IEEE J Biomed Health Inform 20(2):460–468
57. Padmavathia K, Ramakrishna KS (2015) Classification of ECG signal during Atrial Fibrillation using Autoregressive modelling. Int Conf Inf Commun Technol 46:53–59
58. Huang ML, Wu YS (2020) Classification of atrial fibrillation and normal sinus rhythm based on convolutional neural network. Biomed Eng Lett 10:183–193
59. Diker A, Avci E, Tanyildizi E, Gedikpinar M (2020) A novel ECG signal classification method using DEA-ELM. Med Hypothesis 136:109515. https://doi.org/10.1016/j.mehy.2019.109515
60. Wang Q, Mao YW, Ren L, Li Z, Liu H (2020) Automatic classification of ECG data quality for each channel. IEEE Access 8:196094–196101
61. Greenwald SD (1990) Improved detection and classification of arrhythmias in noise-corrupted electrocardiograms using contextual information. Ph.D. thesis, Harvard-MIT Division of Health Sciences and Technology. MIT-BIH Supra-ventricular Arrhythmia Database (SVDB). https://www.physionet.org/, https://physionet.org/content/svdb/1.0.0/
62. Moody GB, Muldrow WE, Mark RG (1984) A noise stress test for arrhythmia detectors. Comput Cardiol 11:381–384. The MIT-BIH NST https://www.physionet.org/, https://physionet.org/content/nstdb/1.0.0/
63. ECG dataset obtained from HCM patients and from control subjects, ICD patients "Hypertrophic cardiomyopathy," Mayo Clinic. [Online]. http://www.mayoclinic.com/health/medical/IM00586. [Accessed 20 Apr 2015]
64. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 101(23):e215–e220
65. Romdhane TF, Alhichri H, Ouni R, Atri M (2020) Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss. Comput. Biol. Med. 123:103866. https://doi.org/10.1016/j.compbiomed.2020.103866
66. Wik L et al (2005) 1631 3-s ECG segments with clinical rhythm annotations Quality of cardiopulmonary resuscitation during out-of hospital cardiac arrest. JAMA 293(3):299–304
67. Albrecht P (1983) S-T segment characterization for long-term automated ECG analysis. M.S. thesis, MIT Dept. of Electrical Engineering and Computer Science. MIT-BIH ST Change (MIT-BIHSTC): https://www.physionet.org/
68. Lichman M (2013) UCI Machine Learning Repository, UCI arrhythmia dataset. School of Information and Computer Sciences, University of California, Irvine
69. Zhang JW, Wang LP, Liu X, Zhu HH, Dong J (2010) Chinese cardiovascular disease database (ccdd) and its management tool. In: 2010 IEEE International Conference on BioInformatics and BioEngineering (BIBE), pp 66–72. CCDD dataset. https://www.physionet.org/
70. Feng N, Xu S, Liang Y, Liu K (2019) A probabilistic process neural network and its application in ECG classification. IEEE Access 7:50431–50439
71. Clifford GD, Liu C, Moody B, Li-wei HL, Silva I, Li Q, Johnson AE, Mark RG (2017) AF classification from a short single lead ECG recording: the PhysioNet computing in cardiology challenge 2017. Computing in cardiology (CinC), vol 44. A fib dataset. https://www.physionet.org/
72. Kung BH, Hu PY, Huang CC, Lee CC, Yao CY, Kuan CH (2021) An efficient ECG classification system using resource-saving architecture and random forest. IEEE J Biomed Health Inform 25(6):1904–1914
73. Sharma M, Tan RS, Acharya UR (2019) Automated heartbeat classification and detection of arrhythmia using optimal orthogonal wavelet filters. Inst Infrastruct Technol Res Manag 16:100221. https://doi.org/10.1016/j.imu.2019.100221
74. Huang J, Chen B, Yao B, He W (2019) ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network. IEEE Access 7:92871–92880
75. Sun Z, Wang C, Zhao Y, Yan C (2020) Multi-label ECG signal classification based on ensemble classifier. IEEE Access 8:117986–117996
76. Yang H, Wei Z (2020) Arrhythmia recognition and classification using combined parametric and visual pattern features of ECG morphology. IEEE Access 8:47103–47117
77. Lu J, Liu D, Liu Z, Cheng X, Wei L, Zhang C, Zou X, Liu B (2021) Efficient hardware architecture of convolutional neural network for ECG classification in wearable healthcare device. IEEE Trans Circuits Syst 68(7):2976–2985
78. Yang J, Yan R (2021) A multidimensional feature extraction and selection method for ECG arrhythmias classification. IEEE Sens J 21(13):14180–14190
79. Yang H, Wei Z (2021) A novel approach for heart ventricular and atrial abnormalities detection via an ensemble classification algorithm based on ECG morphological features. IEEE Access 9:54757–54774
80. Ganguly B, Ghosal A, Das A, Das D, Chatterjee D, Rakshit D (2021) Automated detection and classification of arrhythmia from ECG signals using feature induced long short-term memory network. IEEE Sensors Lett 4(8):1–4
81. Essa E, Xie X (2021) An ensemble of deep learning-based multi-model for ECG heartbeats arrhythmia classification. IEEE Access 9:103452–103464
82. Mahmud T, Fattah SA, Saquib M (2021) DeepArrNet: an efficient deep CNN architecture for automatic arrhythmia detection and classification from denoised ECG beats. IEEE Access 8:104788–104800

83. Nurmaini S, Darmawahyuni A, Rachmatullah MN, Effendi J, Sapitri AI, Firdaus F, Tutuko B (2021) Beat-to-beat electrocardiogram waveform classification based on a stacked convolutional and bidirectional long short-term memory. IEEE Access 9:92600–92613
84. Houssein EH, Abdelminaam DS, Ibrahim IE, Hassaballah M, Wazery YM (2021) A hybrid heartbeats classification approach based on marine predators algorithm and convolution neural networks. IEEE Access 9:86194–86206
85. Hammad M, Iliyasu AM, Subasi A, Ho ESL, Abd El-Latif AA (2021) A Multi-tier deep learning model for arrhythmia detection. IEEE Trans Instrum Meas 70:1–9
86. Li J, Wang G, Chen M, Ding Z, Yang H (2021) Mixup asymmetric tri-training for heartbeat classification under domain shift. IEEE Signal Process Lett 28:718–722
87. Ahmad Z, Tabassum A, Guan L, Khan NM (2021) ECG heartbeat classification using multimodal fusion. IEEE Access 9:100615–100626
88. Kumar A, Ranganatham R, Komaragiri R, Kumar M (2018) Efficient QRS complex detection algorithm based on Fast Fourier transform. Biomed Eng Lett 9(1):145–151
89. Kumar A, Kumar M, Komaragiri R (2019) Design of a biorthogonal wavelet transform based R-peak detection and data compression scheme for implantable cardiac pacemaker systems. J Med Syst 42(6):102
90. Tripathi PM, Kumar A, Komaragiri R, Kumar M (2022) Watermarking of ECG signals compressed using Fourier decomposition method. Multimed Tools Appl 81:19543–19557
91. Zahid MU, Kiranyaz S, Ince T, Devecioglu OC, ChowdhuryKhandakarTahirGabbouj MEHAAM (2022) Robust R-peak detection in low-quality holter ECGs using 1D convolutional neural network. IEEE Trans Biomed Eng 69:119–128
92. Nayak C, Saha SK, Kar R, Mandal D (2019) An efficient and robust digital fractional order differentiator based ECG pre-processor design for QRS detection. IEEE Trans Biomed Circuits Syst. https://doi.org/10.1109/TBCAS.2019.2916676
93. Kumar A, Ranganatham R, Kumar M, Komaragiri R (2021) Hardware emulation of a biorthogonal wavelet transform-based heart rate monitoring device. IEEE Sens J 21(4):5271–5281
94. Berwal D, Kumar A, Kumar Y (2018) Design of high performance QRS complex detector for wearable healthcare devices using biorthogonal spline wavelet transform. ISA Trans 81:222–230
95. Kumar A, Komaragiri R, Kumar M (2019) Design of efficient fractional operator for ECG signal detection in implantable cardiac pacemaker systems. Int J Circuit Theory Appl 47:1459–1476. https://doi.org/10.1002/cta.2667
96. Kumar A, Komaragiri R, Kumar M (2018) Heart rate monitoring and therapeutic devices: a wavelet transform based approach for the modeling and classification of congestive heart failure. ISA Trans 79:239–250
97. Kumar A, Komaragiri R, Kumar M (2019) Time–frequency localization using three-tap biorthogonal wavelet filter bank for electrocardiogram compressions. Biomed Eng Lett 9:407–411
98. Pandey SK, Janghel RR, Vani V (2020) Patient-specific machine learning models for ECG signal classification. Int Conf Comput Intell Data Sci 167:2181–2190
99. Jagtap SK (2013) Digital filter approach for ECG in signal processing. In: Proceedings of international conference on advances in computing, pp 1075–1082
100. Lee Y, Hwang D (2018) Periodicity-based nonlocal-means denoising method for electrocardiography in low SNR non-white noisy conditions. Biomed Signal Process Control 39:284–293
101. Ajdaraga E, Gusev M (2017) Analysis of sampling frequency and resolution in ECG signals. 25th Telecommunication Forum (TELFOR), pp. 1–4. https://doi.org/10.1109/TELFOR.2017.8249438
102. Zhang YT, Zheng Y, Lin WH, Zhang HY, Zhou XL (2013) Challenges and opportunities in cardiovascular health informatics. IEEE Trans Biomed Eng 60(3):633–642

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

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

iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.43227

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

A Waste Segregation Bin

Saurav Kumar[1], Vartika Kaushik[2], Annoy Roy[3]
[1]*Student B.Tech, Chemical Engineering, Delhi Technological University, Delhi, India*
[2]*Student B.Tech, Production and Industrial Engineering, Delhi Technological University, Delhi, India4*
[3]*Student B.Tech, Chemical Engineering, Delhi Technological University, Delhi, India*

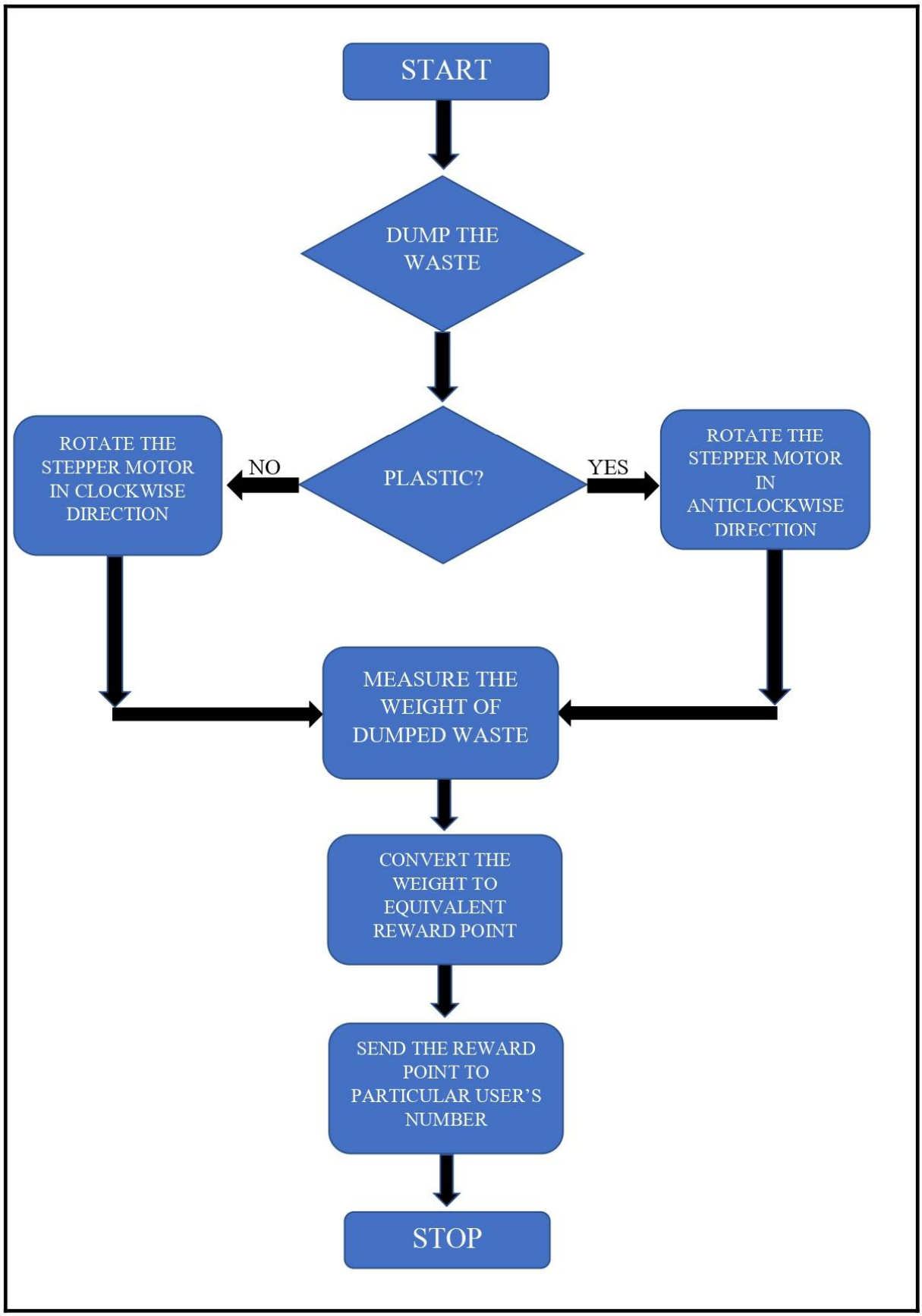
Abstract: ***The primary idea of this project is to create a waste segregation bin which is equipped with a number of sensors to effectively sense the nature of waste and place it in the correct bin. We are also using a LIDAR system so that it can traverse autonomously from house to house in order to collect the waste sample. The system is also equipped with a RFID module so that it can reward the person who uses it.***
The waste bin would also be attached with an ultrasound sensor connected to a web application using NodeMCU so that the authorities can monitor the level of waste present in the bin.
This would save fuel for them and it would also help the companies requiring those waste products to know beforehand how much waste they will get on a particular day.
The companies will pay the authorities some amount for the waste product. The authorities will keep some amount for maintaining the equipment and building more such items and pay the rest to the persons who provide the waste as the data is already available through the RFID card reader.
It would be a benefit for the companies as they are getting the waste at a lower cost, the citizens would get a reward for throwing the waste in the proper bin. We are also planning to use aluminum instead of plastic as it is more durable so less cost of maintenance.
Keywords: Autonomous traversal, Solidworks designing, Electrical,

I. INTRODUCTION

The primary idea of this project is to create a waste segregation bin which is equipped with a number of sensors to effectively sense the nature of the waste and place it in the correct bin.
The system is also equipped with a RFID Module so that it can reward the person who uses it. The waste bin will also be attached with an ultrasound sensor connected to a web application using NodeMCU so that the authorities can monitor the level of waste present in the bin.
This would save fuel for them and it would also help the companies requiring those waste products to know beforehand how much waste they will get on a particular day.
The companies will pay the authorities some amount for the waste product. The bin can traverse from one location to the other through Bluetooth. So that it can go to the household to collect the waste.
The local government and/or environmental NGOs the project for its successful implementation requires a large-scale association of these NGOs and/or governmental organizations and the company who required this trash product for their further manufacturing. When we implement our project on a large scale, we need monetary support either from the government or material support from the aluminum industries, industries manufacturing PCBs, RPLidar, Jetson, motors with encoders, ubiquity, NodeMCU, Arduino and a plethora of other sensors like inductive proximity sensor, inductive capacitive sensors.
This particular project when implemented will require a lot of support from the local government or the local municipal authorities. It will be local municipality who would monitor the level of waste in the bin and carry it to the industries who are willing to buy the waste product at a particular price.

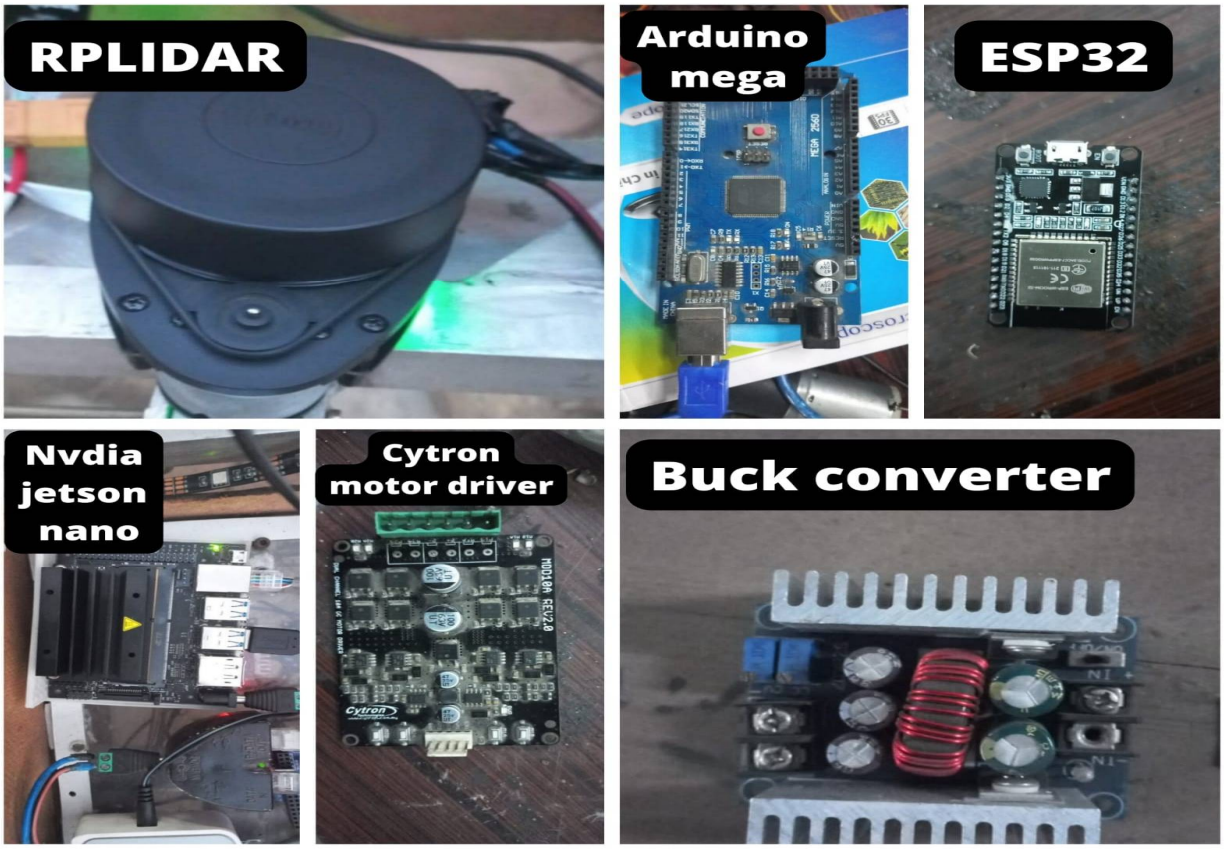
©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com



©IJRASET: All Rights are Reserved

II. COMPONENTS USED



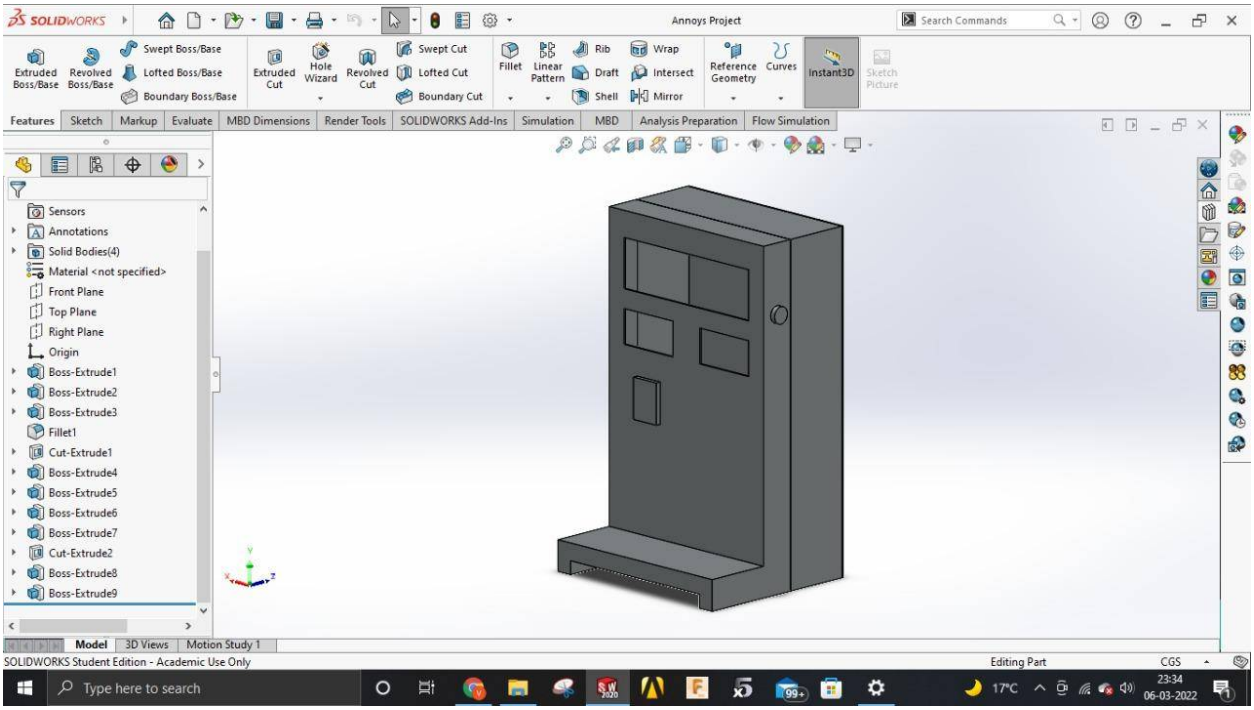
1) Arduino board - The Arduino Uno is a microcontroller board based on the ATmega328 microcontroller from Atmel. It contains 14 digital input output pins, including 6 PWM output pins and 6 analogue input pins. The Arduino Uno can be fueled either by USB or by an external power supply.
2) NodeMCU - NodeMCU is an open source IoT platform with a low cost. It came with firmware that ran on Espressif Systems' ESP8266 Wi-Fi SoC and hardware that was based on the ESP-12 module at first.
3) LIDAR - (Light Detection and Ranging) is a remote sensing technique that measures ranges (varying distances) to the Earth using light in the form of a pulsed laser. When these light pulses are integrated with additional data collected by the aerial system, exact three-dimensional information about the Earth's shape and surface properties is generated.
4) Inductive proximity sensors -An inductive proximity sensor uses electromagnetic radiation to identify metal targets without requiring contact. An inductive proximity sensor's sensing range varies depending on the type of metal detected.
5) Capacitive proximity sensor - Changes in capacitance between the detecting object and the sensor are detected by capacitive proximity sensors. Capacitive proximity sensors work by detecting a change in the capacitance read by the sensor, as the name suggests.
6) Jetson - Makers, learners, and embedded developers now have access to the potential of current AI. The NVIDIA® Jetson NanoTM Developer Kit is a compact but powerful computer that allows you to run multiple neural networks in parallel for image classification, object recognition, segmentation, and speech processing. All of this is contained within a user-friendly platform that consumes as little as 5 watts.
7) Lidar - Lidar is a range-finding technique that involves using a laser to target an item or a surface and measuring the time it takes for the reflected light to return to the receiver. By altering the wavelength of light, it can also be used to create computerised 3-D renderings of locations on the Earth's surface and ocean bottom.
8) Aluminium Sheets

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

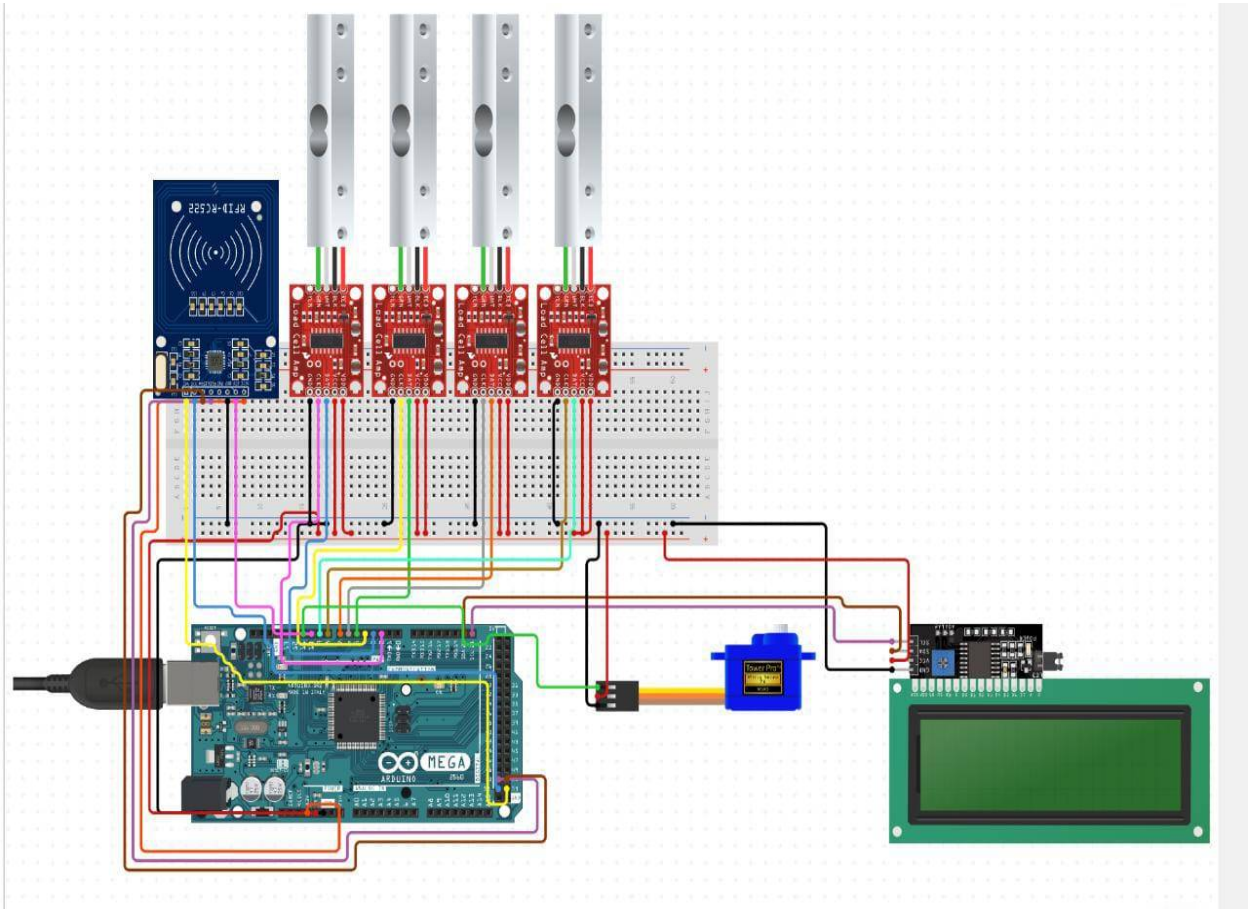
International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

III. DESIGN

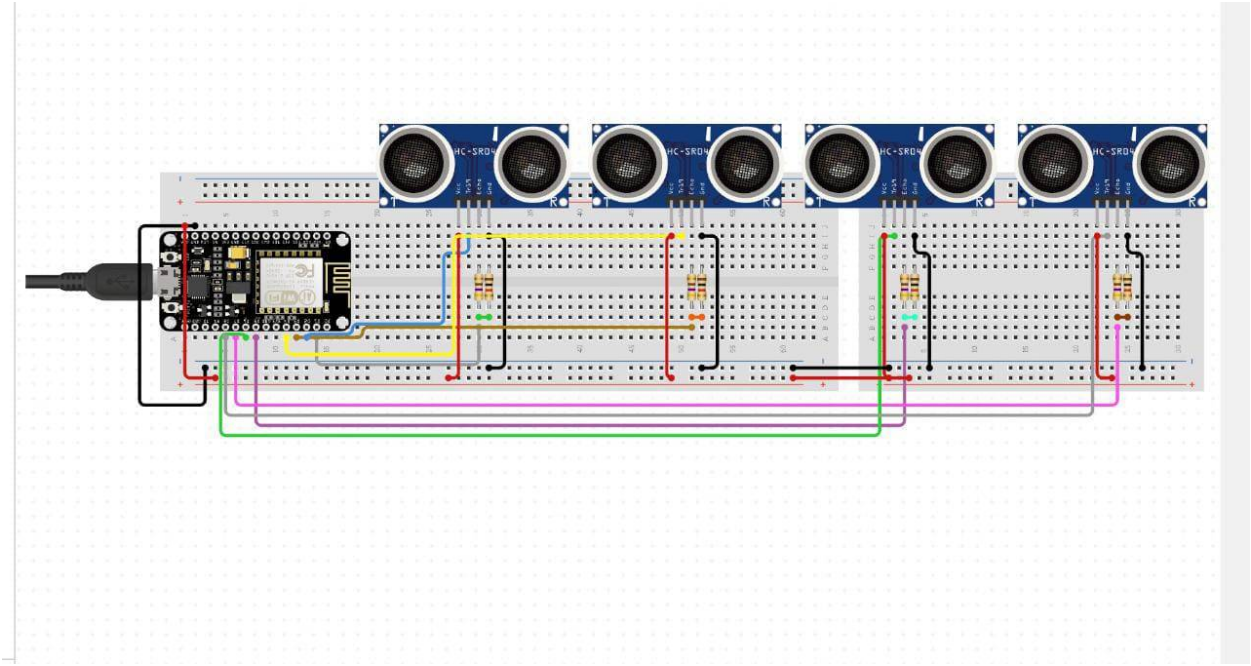
We have designed the dustbin using solidworks designing.



The circuit designing



©IJRASET: All Rights are Reserved



IV. RESULT

Our segregation bin is capable of traversing autonomously from one location to the other. Using a plethora of sensors we are able to achieve about 60 percent accuracy. We are planning to integrate it with computer vision to achieve better results.

V. CONCLUSION

Approaching the correct stakeholder will be the key. After that has taken place, the team shall convince the stakeholder of the magnitude of the problem and the various issues that haphazard collection of waste causes. We would have to explain that, there are both benefits for the person disposing off the waste (who shall get monetary benefits) as well as the environment shall remain clean. We will also have to communicate clearly with the stakeholders on how the industries buying the back the waste shall be benefiting from it. The Stakeholders will have to be convinced that there is absolutely no competition in this field, and the revenue model in paying back the waste thrower is the first of a kind. The raw material used in the dustbin shall be durable and long-lasting helping save money on maintenance of the Dustbin. Lastly, we will have to make people aware of this new trash barrel being installed in their localities. If we overcome the above-mentioned steps, our final step would be the manufacturing and installation of the dustbins at suitable locations.

REFERENCES

[1] Boonrasri, S., Sae-Oui, P., & Rachtanapun, P. (2020). Chitosan and natural rubber latex biocomposite prepared by incorporating negatively charged chitosan dispersion. Molecules, 25(12). https://doi.org/10.3390/molecules25122777

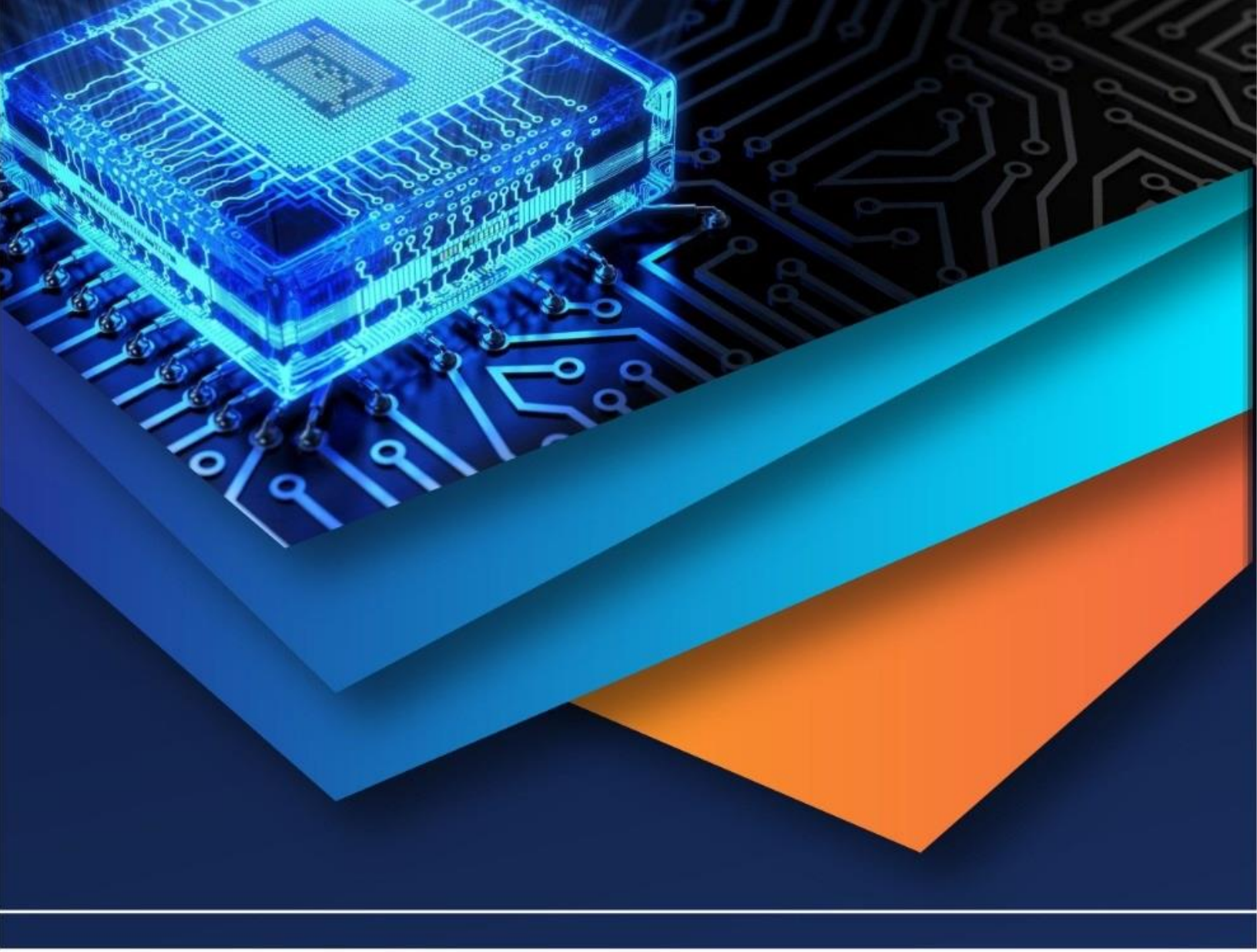
[2] Egbujuo, W. O., Anyanwu, P. I., & Obasi, H. C. (2020). Utilization of chitin powder as a filler in natural rubber vulcanizates: In comparison with carbon black filler. International Review of Applied Sciences and Engineering, 11(1), 43–51. https://doi.org/10.1556/1848.2020.00006

[3] Jang, S. M., Li, M. C., Lim, J. H., & Cho, U. R. (2013). Study on properties of natural rubber compound using starch as filler. Asian Journal of Chemistry, 25(9), 5221–5225. https://doi.org/10.14233/ajchem.2013.f22

[4] Roy, K., Pongwisuthiruchte, A., Chandra Debnath, S., & Potiyaraj, P. (2021). Application of cellulose as green filler for the development of sustainable rubber technology. Current Research in Green and Sustainable Chemistry, 4(April), 100140. https://doi.org/10.1016/j.crgsc.2021.100140

[5] Thomas, S. K., Parameswaranpillai, J., Krishnasamy, S., Begum, P. M. S., Nandi, D., Siengchin, S., George, J. J., Hameed, N., Salim, N. V., & Sienkiewicz, N. (2021). A comprehensive review on cellulose, chitin, and starch as fillers in natural rubber biocomposites. Carbohydrate Polymer Technologies and Applications, 2, 100095. https://doi.org/10.1016/j.carpta.2021.100095

[6] Younes, I., & Rinaudo, M. (2015). Chitin and chitosan preparation from marine sources. Structure, properties and applications. Marine Drugs, 13(3), 1133–1174. https://doi.org/10.3390/md13031133



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)

2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT) | 978-1-6654-1005-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICICICT54557.2022.9917766

An Adaptive Intelligent Approach to Implement Energy Efficiency Concept in Built Infrastructure

Srijan Singh*, Amisha Srivastava, M. Rizwan and Astitva Kumar
Department of Electrical Engineering
Delhi Technological University, Delhi, India
*srijan137b@gmail.com

***Abstract*—The advancement in technology, rapid economic growth and a snowballing population has led to a significant increase in energy consumption globally. The transformation of conventional grid into smart grid where power generation, transmission and distribution are managed intelligently and through a bi-directional automation system is a massive step towards sustainable development. With this transformation, the energy management system becomes essential in order to maintain the resilience of the smart grid. Buildings alone are the dominant contributors to energy consumption worldwide. This paper comprehensively describes the most significant research made in the field of building energy management systems incorporating Internet of things and other intelligent techniques. Further, an analytical approach to implement building energy management system with its detailed architecture is also presented in order to understand the pattern between energy consumption and user preference along with the future scope of research direction. This paper also presents an overview of an intelligent occupancy based approach to control LED lighting systems without causing any trouble to the user.**

Keywords—Building Energy Management System, Smart Grid, Internet of Things, Intelligent Techniques, Smart Cities

I. INTRODUCTION

According to the International Energy Agency, the demand for energy use in the world is expected to increase by more than two-thirds by the year 2035 [1]. Electricity is at the heart of modern economies and its growing demand considerably poses many challenges to the world's electricity system leading to power quality degradation and network congestion problems. Hence begins the transformation of traditional grid into smart grids defining a new era in the electrical sector as we go from classic one-way management to dynamic two-way management thereby increasing efficiency and saving energy [2].

In order to maintain the resilience of smart grid and ensure the stability between supply and demand, Energy Management System (EMS) is a must for electricity suppliers, system operators as well as end-users. EMS is a collection of computerized tools integrated with home automation, buildings, advanced metering infrastructure, electric vehicles and demand side management to monitor and optimize system performance [3]. The technological and infrastructural components of EMS are largely IoT (Internet of Things) based and hence IoT plays a crucial role in building smart grid. The Internet of Things is expected to be widely used in many industrial sectors for day-to-day operations. The broad deployment of sensors coupled with a wide range of physical objects is expected in the next few years. The concept of smart cities is of great interest for researchers. Traffic lights control, street lighting systems, driverless vehicles; smart grid; building management systems (BMS); environment monitoring and cyber security/surveillance are some of the important features of smart city projects. The smart cities comprise 5 sectors namely energy, water, mobility, buildings and security [4]. The challenge is to integrate the applications of all sectors at one place called Smart Building.

India's energy sector is diversified compared to the rest of the World. It is world's third largest energy consuming country [5]. Electricity consumption in India has been estimated at 921 BU (Billion Units) for 2106-17 and forecasted at 1300 BU and 1743 BU for the years 2021-22, and 2026-27, respectively [6]. Per capita consumption of electricity in India from the year 2015 to 2020 is represented in Fig. 1.The proportional share of commercial and residential energy use has been continuously increasing. Between 2001 to 2005, the floor space in India has doubled. Data from the BEE flyer suggests that the overall constructed area will increase to 35 billion sq. m by 2050. Over 35% of the electrical energy consumption in India is associated with building energy consumption [7]. Worldwide consumption of electricity by residential and commercial buildings is nearly 20% [8]. Reports from BEE also suggest that Commercial buildings floor area will grow up to 1900 million m^2 by 2030. Lighting and Air conditioning amounts up to 80% of electricity usage in a commercial building in India while in a residential building fan and lighting loads are prevalent [9]. The sector wise electricity consumption is represented in the form of pie-chart in Fig. 2. Data from NITI Aayog website indicate that Commercial sector energy consumption was 98227.64 GWh and residential sector energy consumption was 288243.1 GWh in 2019 [10]. A huge amount of power is wasted in built infrastructure due to lack of awareness about efficient use of energy, untrained manpower or just negligence. Since this is not possible manually, therefore automation plays a significant role in saving energy in built infrastructure.

This paper discusses various areas of the energy management system and the advancements that have been made in making the smart grid smarter. There are 5 sections including this section. Second section embraces the noticeable research done in the field of building energy management systems making use of IoT and summarizes some of the advances made in the controlling of LED lighting systems. Third section describes an architecture of building energy management system deployed in the Utilization of Electrical Energy (UEE) lab in Delhi Technological University, Delhi, India. It is a platform to monitor and control the lighting systems and air conditioning system as per requirements. Fourth section carefully identifies some of the complexities and the challenges involved in incorporating the proposed control algorithms for efficient energy management. In the next section, the future work of developing controlling techniques for air-conditioning systems is discussed and finally a conclusion is drawn.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:29:45 UTC from IEEE Xplore. Restrictions apply.

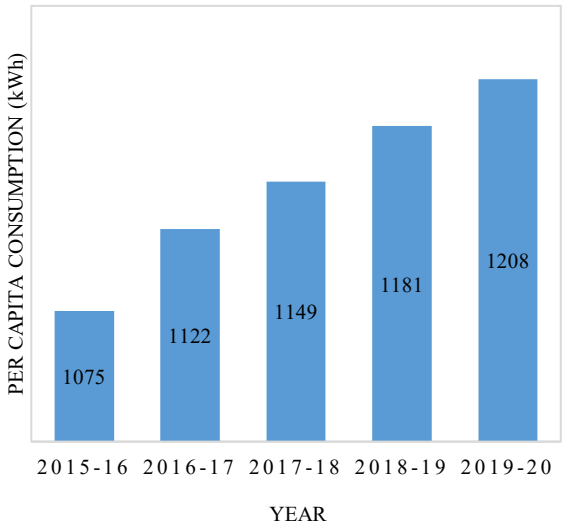


Fig. 1 Per capita consumption of electricity

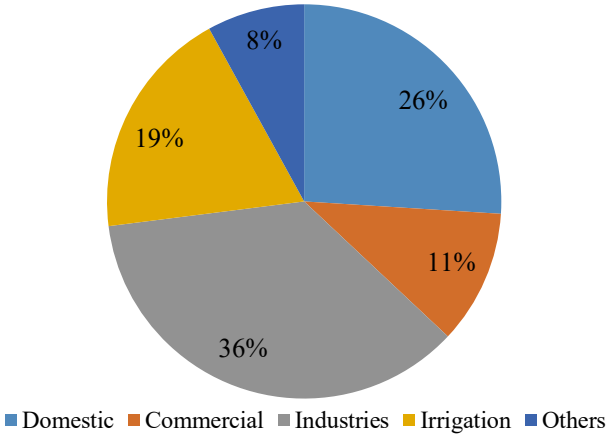


Fig. 2 Sector-wise electricity consumption in India

II. Literature Review

In a commercial/office space, the areas associated with maximum energy consumption are cooling/heating, ventilation, refrigeration and electronics items like computers. This fast growing trend of energy consumption is worrying but at the same time provides an opportunity to look for solutions to bring down the energy levels. Hence, existing and new solutions must be incorporated to the Building Management Systems (BMS). Sensors are now available at a lower price, user friendly open-source platforms are available; thus integration of IoT devices is pacing up in smart homes/ building management systems.

BMS is a sophisticated platform used to monitor and control a building's load demand. It is used to manage and enhance the efficiency of the pre-existing loads and install new loads with less power consumption. A BMS platform communicates with the hardware and performs tasks of cooling, heating, lighting, ventilation, video surveillance, switching on fire alarms etc. The sensors take the input from the surroundings. This input can be temperature, humidity, CO2 levels, image, sound, smoke, illuminance, room occupancy etc. The sensors then transmit the collected data from the surroundings to the controllers. The controllers then form a decision based upon the inputs they receive. The output devices then implement the commands given by the controllers. The queries and results are shown on the user's Graphical User Interface. The parameters like voltage, current, power, power factor are sensed in real time and are up for display on the BMS platform.

Santiago et. al in [11] have discussed about BIM (Building Information Modeling) methodology to obtain energy model of the buildings. The simulation platform of BIM software has shown considerable reduction in energy consumption. Energy savings of 61.09 kWh/m^2/ year for Air-conditioning systems were achieved and 26.8 kWh/m^2/year of energy was conserved from LED lighting systems.

Jamil and Mittal in [12] have mentioned about building energy management softwares. Some of the examples are Freedomotic, Open Remote and openHAB, these softwares have a limitation that they cannot implement Demand Response. The authors also mentioned about BEMOSS, an open source software which can monitor and control loads in small (less than 5000 sq.ft.) and medium (5000-50,000 sq.ft.) sized buildings. [13] Extensive work on how to integrate IoT devices with BEMOSS (Building Energy Management Open Source Software) is explained. Some of the major loads like HVAC (Heating, Ventilation, Air Conditioning), Plug loads, Dimmable blasts can be installed and controlled via BEMOSS platform. BEMOSS supports prevalent communication technologies like Ethernet (IEEE 802.3), Serial (RS-485), ZigBee (IEEE 802.15.4) and Wi-Fi (IEEE 802.11); and data exchange protocols like BACnet, Modbus, Web, ZigBee API, Open-ADR and Smart Energy (SE) protocols.

In [14], authors mention an important point of establishing a coordinated energy management system. The HEMS (Home Energy Management Systems) operates by scheduling loads, by establishing a two way communication with the grid and other customers to prevent any power outage/ blackouts or overloading. The coordination can be either distributed or centralized. Researchers also explain that the household energy efficiency can be reduced through consumption reduction or consumption shifting. Consumption reduction refers to energy conscious building designs, usage of devices with less power consumption but with higher efficiency. Consumption shifting means to defer certain loads during the peak demand. Consumption shifting is based on the methods which can find optimal operational timing for appliances.

Building energy consumption can be carried out by using predictive models. Estimations to find consumption for different purposes like cooling, heating, lighting are mentioned in [15]. Methods which can predict micro-generation for solar and wind are described in [16]-[18]. The following can be seen as a forecasting problem which can be solved through machine learning algorithms, [17], [19], [20]. Mathematical Optimization, Machine Learning, ANN based techniques are mentioned which can be used for load scheduling [21].

LED Lighting systems- Changes made in the lighting sector is mainly due to the eco-friendly nature of LED, high-energy efficiency and long life. However, installation cost comes across as a deterrent and hence the large scale of deployment of LED is a big challenge. As a result, numerous research projects have been conducted to find the fix for economic feasibility of LED lighting. Some of the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:29:45 UTC from IEEE Xplore. Restrictions apply.

TABLE I. SUMMARY OF SELECTED LITERATURES

Ref.	Objective of the study	Method	Future Scope of work	Area of Usage
[13]	To give a comprehensive study of the software BEMOSS	BEMOSS is an open-source software designed for control and monitoring of hardware devices. It supports wireless technologies like Wi-Fi, Zigbee and wired technologies like Ethernet, Serial RS-485.	BEMOSS is specifically designed for the built infrastructure in USA, can be further designed for commercial buildings in other countries	Small (<5000 sq.ft.) and Medium (less than 50,000 sq.ft.) sized commercial buildings
[23]	Minimize power consumptions based on occupancy levels in different zones, while considering daylight conditions	Ultrasound sensors detect the movement accurately and then feed this information as input to the control algorithm, which serves the purpose of dimming.	Testing Conditions based on different occupancy levels and its impact on energy savings	Proposed system can work for office/residential spaces
[24]	Develop Lighting Control System to meet user requirements and save energy	Using wireless sensors network different light intensities are gathered. Algorithms to control illuminance and devices take care of user preferences.	Distribution of luminaires and energy wastage needs to be checked	Indoor complexes of residential/ office spaces
[25]	To study energy efficiency patterns without a comprise in user's activity patterns	ESLiCoS method used here consists of 6 parameters to control lighting patterns.	Algorithms that can control lights according to the nature of work is an area that can be further explored	Corridors, Reading Rooms, Washrooms
[26]	Emphasis on adjusting brightness of lights according to the user's preferences.	Solid State Lighting control implemented on wireless sensor network shows desirable results of preferred illuminations at desired places.	Algorithms considering multiple user preferences in a single zone can be studied and implemented further.	Indoor Lighting Systems for Residential and Office spaces
[27]	Reduce domestic energy usage and utility bills, along with ensuring an optimal task scheduling and a thermal comfort zone for the inhabitants.	A Multi-Objective Mixed Integer Nonlinear Programming (MO-MINLP) model is developed for efficient energy use in a home considering energy reduction, user's convenience rate and thermal comfort level (TCL).	Improved algorithm taking account of more real-time data from smart homes or residential micro-grids	Small, medium sized houses, residential micro-grid
[28]	Design an efficient airconditioning control system that decides the number of units to be operated based on the tenancy of a building	An automatic system using open source micro-controller (Arduino) was used making use of sensors to detect objects. Accordingly, programming is done to turn on/off the air conditioner.	More intelligent motion detectors such as 3D cameras can be used to authenticate human presence. Can be used to control multiple devices simultaneously.	Schools, colleges, offices, homes

control algorithms have been in mentioned in [22]-[24]. Control algorithm for device control and illuminance control are mentioned in [22], which is based on wireless sensor network (WSN). Pandhripande and Caicedo have shown in [23] that LED lighting can be controlled based on the location of the occupants and taking account of the daylight conditions. The illuminance levels are adjusted according to user's activities. The concept of neural network based approach to control LED lighting systems is mentioned in [24]. It takes account of the relationship between the dimming light levels and table illuminance. In [25] Byun and Shin have introduced ESLiCoS (Energy-Saving Lighting Control System) method based on user preference for controlling lighting systems. The proposed system adjusts the brightness of LED's using six control parameters thereby reducing energy consumption by 43%. As of now, control of lighting systems using is a great topic of interest among researchers. Parameters such as movement of people and daylight conditions can be used to control the intensity of lights and set the illuminance according to the user's preference.

III. PROPOSED BUILDING ENERGY MANAGEMENT SYSTEM ARCHITECTURE

Existing LED systems are known to have reduced the energy consumption but adaptive nature according to occupancy is a concern. Hence, the existing technologies are mostly used in walkways, corridors, staircases. Office spaces having committee rooms, conference rooms and other areas like classrooms need to have occupancy based lighting system. In this section, the area under consideration is the UEE Lab at Delhi Technological University, Delhi, India. The lab having dimensions of 40 x 40 ft. is divided into 8 zones and each zone has 32 LED lights, having ratings of 18W and 40W.

Therefore, the challenge is to control 32 lights in 8 different zones at different points of time with varying occupancy. This paper not only proposes a concern for energy reduction but also proposes user-satisfaction along with energy reduction. The framework for IoT based smart building energy management system for a location is shown in Fig. 4.

TABLE II. TYPES OF LOADS IN UEE LAB

S.No.	Loads and Ratings		
	Load Name	*Quantity*	*Rating*
1.	LED Light	24	40 W
2.	LED Light	8	18 W
3.	AC Fans	16	60 W
4.	Air Conditioners (Split)	3	1600W
5.	Air Conditioners (Window)	3	1680 W

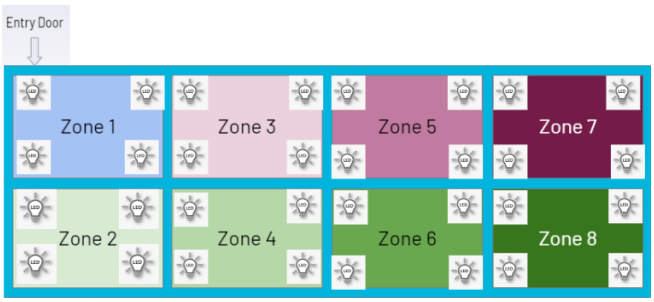


Fig. 3 Zonal Division of Laboratory

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:29:45 UTC from IEEE Xplore. Restrictions apply.

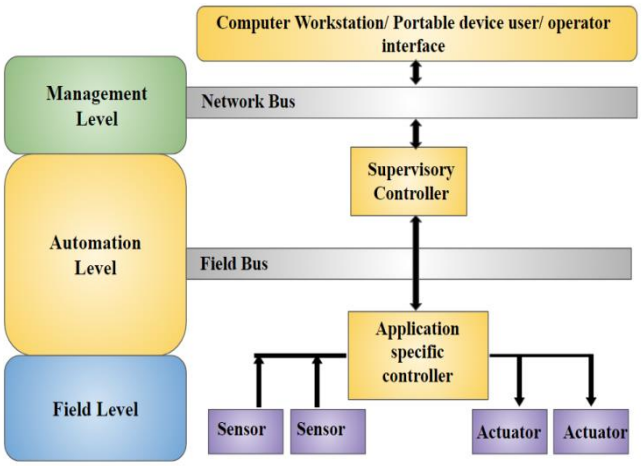


Fig. 4 Framework for IoT based smart building energy management system for a location

The sensor labeled in Fig. 5 is the occupancy sensor. The sensor gathers the input from the surroundings i.e. it senses the movement of people around. This data is provided to the Node-MCU. The Node-MCU when configured with the arduino-software displays the output on serial monitor. When the movement is detected, the output is 1 and no movement corresponds to zero. The movement can be as simple as a wave of hand. To distinguish the proposed system from an ordinary on/off system, this paper takes account of the user-preference. The illumination of the lights will be fixed at a minimum level for e.g. 40% at places of high occupancy and this level can be adjusted to 25% at places of minimum occupancy. A counter of 30 seconds will send a signal to the controller to reduce the brightness to zero if no motion is sensed during that interval. Similarly, when the light is operating at maximum illuminance, the counter will wait for 30s to detect any movement, if motion is sensed during that interval the light brightness will begin to dim, as shown in Fig. 6. The BMS platform that will be developed under this project will show the results on the API (Application Programming Interface) platform. Illuminance levels, power consumption, hourly average are some of the parameters of interest. The objective of the study is to bring down energy levels without any discomfort to the laboratory personnel. Upon successful integration of the sensors with Wi-Fi and BMS platform graphs showing the energy reduction trends will be shown.

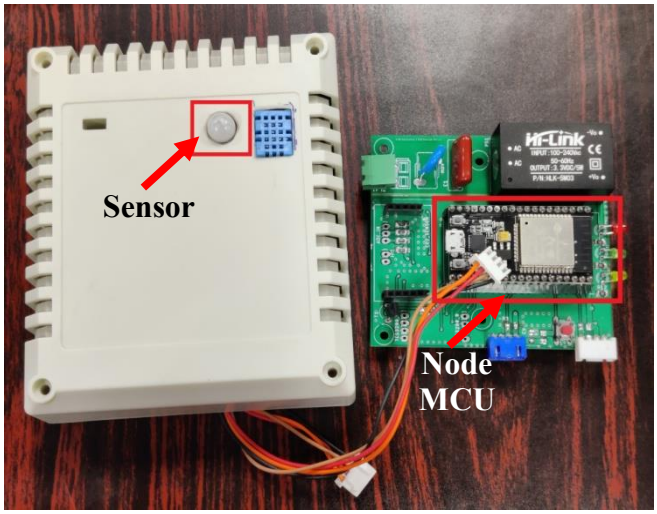


Fig 5. Hardware for Controlling Lighting System

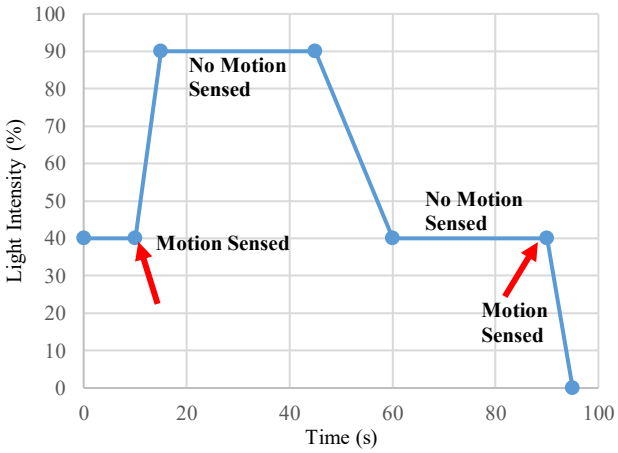


Fig 6. Proposed lighting timing system

IV. Challenges

Building Energy Management is a broad concept and managing the energy and additional necessities in buildings efficiently and intelligently can have substantial benefits but there are certain limitations that cannot be disregarded. The initial cost for designing and installation of intelligent technologies is quite high and since BEMS consists of sensing and controlling devices, their updating and maintenance cost is even higher. Sometimes even acquiring data using cloud-based analytics making use of high-end software is also difficult.

BEMS is a great channel for the implementation of IoT technology and leading specialists are constantly innovating to bridge the gap in compatibility between software and hardware components of IoT, which leaves plenty of potential entry points for attacks from cyber criminals, i.e., hackers. This requires future designs to be more cryptographically secured and advanced.

Existing technologies related lighting control mainly focuses on energy reduction. However, energy reduction at the expense of convenience is not desirable. The complexity levels of an user activity will further determine the minimum and maximum illuminance levels. Activities like reading, studying would require different illuminance levels, frequently changing brightness is not desirable. Thus, adaptive lighting control systems can be placed in certain sections where the illuminance does not falls or rises suddenly.

V. Conclusion

The ultimate objective of the paper is to present an extensive literature survey on building energy management system and identifying the research gaps. Energy can be effectively optimized by honest business practices and the lucrative incentives provided by the state/central government. According to the Energy Conservation Building Code (ECBC), the Energy Performance Index (EPI) of Commercial buildings in India needs to be brought down to 100-150 kWh/m^2/year. The facts and figures clearly describe the global energy scenario.

The proposed architecture of BEMS implemented in built infrastructure i.e. UEE Laboratory in Delhi Technological University, Delhi, India is a step taken towards sustainable development of the technology

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:29:45 UTC from IEEE Xplore. Restrictions apply.

employed in this work. Future works aim at incorporating more tools and technologies for end-to-end management of such complex systems. The future scope of the present work is to develop control algorithms coupled with IoT devices for optimal use of air conditioning loads in context of UEE lab, Delhi Technological University.

ACKNOWLEDGEMENT

The proposed research was carried out as a part of sponsored research project funded by SERB, vide File No. CRG/2018/000291.

REFERENCES

[1] International Energy Agency, "India Energy Outlook 2021", Paris, France, 2021.

[2] S. V. Sreedevi, P. Prasannan, K. Jiju, and I. J. Indu Lekshmi, "Development of indigenous smart energy meter adhering Indian standards for smart grid," in *2020 IEEE International Conference on Power Electronics, Smart Grid and Renewable Energy (PESGRE2020)*, 2020.

[3] K. Mahmud, D. Soetanto, and G. E. Town, "5.6 Energy Management Softwares and Tools," in *Comprehensive Energy Systems*, Elsevier, 2018, pp. 202–257.

[4] D. Minoli, K. Sohraby, and B. Occhiogrosso, "IoT considerations, requirements, and architectures for smart buildings—energy optimization and next-generation building management systems," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 269–283, 2017.

[5] bp, "Statistical Review of World Energy2021," London, UK, 2021.

[6] Central Electricity Authority, "19thElectric Power Survey Report," New Delhi, India, 2020.

[7] Bureau of Energy Efficiency, "Energy benchmarks for Commercial Buildings," New Delhi, India, 2020.

[8] Energy Information Administration, "International Energy Outlook 2019," Washington DC, USA, 2019.

[9] S. Seth, "Energy Effeciency Initiatives in Commercial Buildings," Bureau of Energy Effeciency, New Delhi, India, 2012.

[10] NITI Aayog, "India Energy Dashboards, 2019" New Delhi, India, 2019.

[11] F. J. Montiel-Santiago, M. J. Hermoso-Orzáez, and J. Terrados-Cepeda, "Sustainability and Energy Efficiency: BIM 6D. Study of the BIM Methodology Applied to Hospital Buildings. Value of Interior Lighting and Daylight in Energy Simulation," *Sustainability*, vol. 12, no. 14, p. 5731, Jul. 2020, doi: 10.3390/su12145731

[12] M. Jamil and S. Mittal, "Building Energy Management System: A Review," in *2017 14th IEEE India Council International Conference (INDICON)*, 2017, pp. 1–6.

[13] W. Khamphanchai, M. Pipattanasomporn, M. Kuzlu, and S. Rahman, "An agent-based open source platform for building energy management," in *2015 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA)*, 2015, pp. 1–6.

[14] J. Leitão, P. Gil, B. Ribeiro and A. Cardoso, "A Survey on Home Energy Management," in *IEEE Access*, vol. 8, pp. 5699-5722, 2020.

[15] T. Ahmad, H. Chen, Y. Guo, and J. Wang, "A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review," *Energy Build.*, vol. 165, pp. 301–320, 2018.

[16] D. W. van der Meer, J. Widén, and J. Munkhammar, "Review on probabilistic forecasting of photovoltaic power production and electricity consumption," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 1484–1512, 2018.

[17] D. W. van der Meer, M. Shepero, A. Svensson, J. Widén, and J. Munkhammar, "Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using Gaussian Processes," *Appl. Energy*, vol. 213, pp. 195–207, 2018.

[18] L. Chinmoy, S. Iniyan, and R. Goic, "Modeling wind power investments, policies and social benefits for deregulated electricity market – A review," *Appl. Energy*, vol. 242, pp. 364–377, 2019.

[19] M. Molina-Solana, M. Ros, M. D. Ruiz, J. Gómez-Romero, and M. J. Martin-Bautista, "Data science for building energy management: A review," *Renew. Sustain. Energy Rev.*, vol. 70, pp. 598–609, 2017.

[20] D. Zhang, X. Han, and C. Deng, ``Review on the research and practice of deep learning and reinforcement learning in smart grids," *CSEE J. Power Energy Syst.*, vol. 4, no. 3, pp. 362_370, Sep. 2018

[21] M. Bourdeau, X. Q. Zhai, E. Nefzaoui, X. Guo, and P. Chatellier, "Modeling and forecasting building energy consumption: A review of data-driven techniques," *Sustain. Cities Soc.*, vol. 48, no. 101533, p. 101533, 2019.

[22] M.-S. Pan, L.-W. Yeh, Y.-A. Chen, Y.-H. Lin, and Y.-C. Tseng, "A WSN-based intelligent light control system considering user activities and profiles," *IEEE Sens. J.*, vol. 8, no. 10, pp. 1710–1721, 2008.

[23] A. Pandharipande and D. Caicedo, "Daylight integrated illumination control of LED systems based on enhanced presence sensing," *Energy Build.*, vol. 43, no. 4, pp. 944–950, Apr. 2011.

[24] Z. Wang and Y. K. Tan, "Illumination control of LED systems based on neural network model and energy optimization algorithm," *EnergyBuild.*, vol. 62, pp. 514–521, Jul. 2013.

[25] J. Byun and T. Shin, "Design and implementation of an energy-saving lighting control system considering user satisfaction," *IEEE trans. consum. electron.*, vol. 64, no. 1, pp. 61–68, 2018.

[26] S. Bhardwaj, T. Ozcelebi, R. Verhoeven, and J. Lukkien, "Smart indoor solid state lighting based on a novel illumination model and implementation," *IEEE Trans. Consum. Electron.*, vol. 57, no. 4, pp. 1612–1621, Nov. 2011.

[27] A. Anvari-Moghaddam, H. Monsef and A. Rahimi-Kian, "Optimal Smart Home Energy Management Considering Energy Saving and a Comfortable Lifestyle," in *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 324-332, Jan. 2015.

[28] A. M. Hashmi, Ibrahim A. Alkabiri, Khaled W. Alnajjar, Hatim H. Alattas, and Talal H. Almutaiti, "Design of a Smart Air Conditioning Controller Based on the Occupancy of a Building," *International Journal of Electrical and Electronic Engineering & Telecommunications*, Vol. 10, No. 3, pp. 160-167, May 2021. Doi: 10.18178/ijeetc.10.3.160-167

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:29:45 UTC from IEEE Xplore. Restrictions apply.

An Effective Multimodal Biometric System Based on Textural Feature Descriptor

Neeru Bala[a,*], Rashmi Gupta[b,], and Anil Kumar[a,***]**

[a] *Amity School of Engineering and Technology, Amity University Gurugram, Haryana, 122413 India*

[b] *Electronics and Communication department, Netaji Subhash University of Technology, East Campus, Delhi, 110031 India*

** e-mail: neeru.panghal1@gmail.com*

*** e-mail: rashmi.gupta@nsut.ac.in*

**** e-mail: akumar2@ggn.amity.edu*

Abstract—In recent years the upsurge of contemporary technical assets results in the augmented prerequisite of precise and robust security systems. The utmost prevailing alternative for the security requirements is the deployment of effective biometric authentication frameworks. The unimodal biometric frameworks have various snags due to which they lag in precision. In this concern, the multimodal biometrics has gained more consideration of investigators as it utilizes the proven and effective strategy to integrate the multiple human traits for authentication. This paper proposes a multimodal biometric framework, which is based upon score level fusion of iris, palmprint, and finger-knuckle-print (FKP). It is a touchless system as these traits can be acquired without any physical contact with the sensors, so this system maintains the hygiene factor, which is the need of the hour in this pandemic time. Rigorous experimentations on three benchmark databases reveal that the proposed multimodal framework is proficient in augmenting the recognition rate in comparison to the unimodal framework as well as prevailing multimodal state of art works on attaining lowest equal error rate (EER) of 0.56% and highest genuine acceptance rate (GAR) of 99.70%.

Keywords: information security, multimodal biometrics, information fusion, score level fusion

DOI: 10.1134/S1054661822030063

1. INTRODUCTION

Biometric recognition is based upon various automated processes of authenticating an individual in verification and identification mode [21]. Various biotic traits such as iris [9, 12], palmprint, earprint, DNA, finger-knuckle-print, finger-vein [7, 53], face [8, 26], and behavioral traits such as voice, gait [39, 48], or handwriting, have been used as traits in biometric frameworks [17]. Biometrics has an extensive variety of application areas such as in national identification card (e.g., Aadhar card), border control, law enforcement, secure buildings, ATM machines, banks, or personal gadgets [47]. Traditional unimodal biometric frameworks use one trait for recognizing an individual [46]. But they suffer from various snags such as dearth of distinctiveness, noisy data non-universality, deplorable error rates, and susceptibility to fraudulent attacks. To mitigate these snags, the use of multimodal biometric frameworks is proven to be an effective technique. A multimodal biometric framework utilizes more than one trait for authenticating an individual. The fusion of traits can be performed on various levels such as (i) sensor level fusion, where traits are fused before feature extraction; (ii) feature level fusion [38], where feature vectors obtained from multiple traits are fused; (iii) rank level fusion, where ranks of different traits are fused [15]; (iv) decision level fusion, where decisions corresponding to different traits are fused to achieve the final decision about recognition; and (v) score level fusion, where the genuine and imposter scores of different traits are fused [2].

A biometric process operates in two levels, enrolment and identification level [20]. At enrolment level biometric traits are acquired from every user and then it is converted into a template by processing it with particular algorithms. The templates so obtained, of each user are stored in the database. At identification level, biometrics sample images are acquired and processed with the same algorithms as at enrolment level, and then these processed sample images are matched by the user's templates kept in the dataset. And then, on the basis of resemblance of sample images with templates, a decision about recognition of an individual is taken. Usually, there are two functioning modes of a biometric system named verification mode and identification mode. In positive recognition, the verification mode is employed such as in the attendance system of an organization where, an employee claims its identity by using PIN, ID card or username; then, the system maps the claimed sample template with the template of that person already stored in the database, so in verification mode one-to-one mapping of tem-

Received October 29, 2021; revised April 21, 2022; accepted April 22, 2022

plates take place. Whereas, in case of identification mode one-to-many mapping takes place as there is no claimed identity, so the sample template of the person to be identified is matched with the whole templates stored in the database. The negative recognition identification mode is usually employed, such as criminal identification at the border of a nation.

In recent years, numerous multimodal biometric systems using different fusion strategies for fusion of different modalities to obtain augmented authentication have been proposed. Some of the recent research works on multimodal biometric systems include the following. Sarangi et al. [45] integrated features of ear and profile face using a kNN classifier based technique. Xiong et al. [60] introduced a strategy based on swarm optimization for selecting features of face and iris to be fused. Cherifi et al. [10] combined ears shape and arm gesture at score level. Weighted Euclidean distance based fusion of palmprint and palm-vein was introduced by Misar and Gharpure. Gupta et al. [16] suggested fusion of iris and fingerprint to generate cancelable biometrics. Ammour et al. [4] proposed integration of facial image and iris at score level.

A general biometric recognition framework comprises four sections such as pre-processing of images, extraction of features, classification and decision stage. The selection of feature extraction technique prominently affects the efficacy of any biometric system. There are several feature extraction techniques. In this work we have employed improved X-Or sum code with curvature Gabor filter (CGF) [6] for feature extraction of iris, palmprint and finger-knuckle-print. Curvature Gabor filter is proficient in extracting orientation as well curvature information present in the images. Moreover, these three traits have been integrated at score level. Highpoints of this research paper are as follows:

- an effective multimodal biometric system to integrate iris, palmprint, and finger-knuckle-print (FKP);
- feature extraction via Improved X-OR sum code (IXSC);
- integration of these traits at score level;
- the experimentation results reveal low equal error rate (EER) and high genuine acceptance rate (GAR).

1.1. Motivation

Due to lack of information, the unimodal biometric framework cannot fulfill the requirements of higher authentication performance. To accomplish the augmented authentication performance, fusion of iris, palmprint, and finger-knuckle-print is an obvious approach. The research motive to fuse iris, palmprint, and finger-knuckle-print biometrics is to employ them all together to achieve superior authentication performance, which may not be accomplished by utilizing them individually. Iris is the annular region existent between pupillary and limbic boundaries [9]. The atypical texture existent in this region makes the iris capable of being used as a biometric trait [11]. Iris has consistently been the choice of researchers with regards to the design of an exceptionally precise and spoof free [21] biometric acknowledgment framework. The iris has consistently been the selection of analysts with regards to the plan of an exceptionally exact [9] and altering free [21] biometric acknowledgment framework. The most common traits such as face and fingerprint are contenders of iris but unlike iris, these do not stay steady throughout the person's lifetime [32]. Furthermore, the iris has least effect of deterioration of exterior surroundings as compared to other existent traits [12]. The iris also possesses the hygiene factor as it is a contactless trait. Undue advances in the image acquisition techniques have reduced the requirement of subject's support while capturing the images for authentication [21]. All these specifics makes the iris a distinctive, versatile, consistent, accurate and contactless biometric trait, so it is an effective choice as an identity measure in biometric authentication.

Whereas, the palmprint is the interior side of human hand that comprises numerous discerning characteristics such as principal lines, geometry-features, ridges, delta points and wrinkles. The palmprint is a profound unique identifier. From the past decade the palmprint has emerged as an effective biometric trait as it provides the unique and large number of features due to its large surface area in comparison to several traits such as fingerprint, which results in enhanced recognition rate [27]. Moreover, palmprint acquirement is easier and cheaper as it is acceptable to most of subjects, and also it remains stable throughout lifetime of a subject in comparison to traits such as face. Henceforth, palmprint recognition framework possesses potential to be used in developing reasonably secure biometric systems [19]. The finger-knuckle-print is an emergent biometric trait. The pattern produced by proximal inter phalangeal (PIP) joint on back region of finger is known as finger-knuckle-print [24]. As FKP concentrates on the interior region of finger, it is a very challenging task to temper it in comparison to fingerprint. Finger-knuckle-print is a novel, unique source of rich information trait which is well appropriate to be utilized in developing a potential biometric framework [37]. Henceforth, all three traits have their own advantages, but due to ailment, if any user cannot provide the samples of any trait, then there will still be an option to access the system. Furthermore, it is a very challenging task to tamper three traits simultaneously. So, we have developed a novel robust multimodal biometric system which uses three traits to authenticate an individual,

The rest of the paper is systematized as follows: in Section 2 we briefly summarize some recent related works in the arena of multimodal biometric framework, Section 3 refers to the proposed approach, Section 4 showcases the experimentations performed to

validate the proposed approach, and Section 5 presents the conclusion.

2. RELATED WORK

To accomplish the augmented authentication rate, fusion of various traits is an obvious approach. Considerable attention has been paid in this context and several methodologies have been proposed. A brief review of some cutting-edge research works on multimodal biometric frameworks is presented here. Thepade and Bhondave [52] proposed a comparative analysis of feature level and score level fusion of iris and palmprint. Rajagopal and Palaniswami [42] proposed fusion of iris and palmprint at feature level. Naderi et al. [34] presented feature level fusion of iris and palmprint to obtain enhanced authentication. They have extracted the feature of these traits using four descriptors such as Log Gabor, DCT, Walsh and Haar, and Laplacian eigenmaps, and singular value decomposition (SVD) to overcome the huge dimensionality issue in feature-level fusion. Mayya and Saii [31] introduced a wavelet transform based score level fusion of iris and palmprint to achieve augmented authentication. A robust multimodal framework based on feature level fusion was suggested by Attallah et al. [5]. More et al. [33] introduced proportional analysis of unimodal and multimodal biometrics system fusing iris and palmprint at score level. Multiresolution based approach to fuse iris and palmprint was suggested by Hariprasath and Santhi [18] with 94.50% recognition rate. An approach to fuse iris and palmprint based on fuzzy classifier was introduced by Alsubari et al. [3] with recognition rate of 98.96%. Zhang et al. [61] fused scores of face and voice for augmenting recognition rate. Jain et al. [22] proposed score level fusion of iris and sclera.

Amritha and Aravinth [55] integrated ECG, fingerprint, and face at score level. Ramachandran and Sankar [43] fused Iris and Palmprint at score level. Abderrahmane et al. [1] combined fingerprint and face at score level. Rane [44] proposed an approach to fuse scores of palmprint and face. Srivastava [50] integrated the scores of retina, finger-vein, and fingerprint. Rahman [41] integrated ECG and fingerprint at score level. Jaswal and Poonia [23] suggested rank level fusion of palmprint and finger-knuckle-print. Devi and Rao [14] combined palmprint and face at decision level. Naji [36] suggested decision level fusion of fingerprint and face. Purohit and Ajmera [40] proposed optimal feature level fusion technique to combine fingerprint, ear, and palmprint. Joseph et al. [25] suggested fusion of fingerprint, iris and palmprint at feature level. Vinothkanna and Wahi [56] combined features of fingerprint, palmprint, and hand-vein. Soviany et al. [49] integrated fingerprint and palmprint at feature level. Srivastava [50] proposed a multimodal biometric framework to fuse face, finger-print, and palmprint at score level. Vyas et al. [57] proposed an accurate feature extraction technique for iris and palmprint.

In the literature various multibiometric strategies for integration of multiple traits at various levels of fusion have been discussed. Feature level fusion and score level of fusion are most prevalent in the literature for fusing the traits but feature level fusion has dimensionality issues. Until now, there are fewer studies on fusion of iris, palmprint, and finger-knuckle-prints for human recognition.

3. PROPOSED APPROACH

In this paper, a novel multimodal recognition framework integrating iris, palmprint, and finger-knuckle-print (FKP) has been proposed. The proposed approach was implemented using improved X-OR sum code [6], in which features of each trait were extracted using curvature Gabor filter and then these traits were fused at score level using the sum and product rule. This section exemplifies the facts of the proposed approach in detail, and a schematic structure of the proposed approach is presented in Fig. 1.

3.1. Preprocessing of Images

In a biometric recognition framework, the preprocessing of images is the substantial step. The objective of this step is to augment the quality of images so that these images can provide a huge amount of vital information and furthermore this information helps in extracting the region of interests (ROIs). In case of iris, initially the eyes are distinguished utilizing the Viola–Jones algorithm [4]. It depends on the rule that they are obscurer than other portions in the face. It discovers the little patches which are darker than their surrounding and are of approximately the same size as that of an eye. After detection, iris images are segmented and normalized to a rectangular strip of size 64×512 pixels using the John Daugman's approach [13]. For extracting ROI of palmprint, characterizing certain reference indicator is important to depict the relative interpretation, rotation and scale between various palmprint images. In this work, ring-small finger and index-middle finger are defined as reference milestones. The y-axis of the ROI is defined as the line fragment between two reference points, and then a square-shaped ROI is extracted using lateral length equals to 3/2 of the y-axis of ROI. Normalization of extracted ROI is done using histogram equalization which results in resized square-shaped strip of dimension 144×144 pixels [35]. For finger-knuckle-print ROI extraction the canny edge detector is applied to finger image, which works on the basis of intensity of pixels. For ROI extraction the density of pixels with high intensity is used. Therefore, a region that consists of typically edge elements and finger symmetry line of

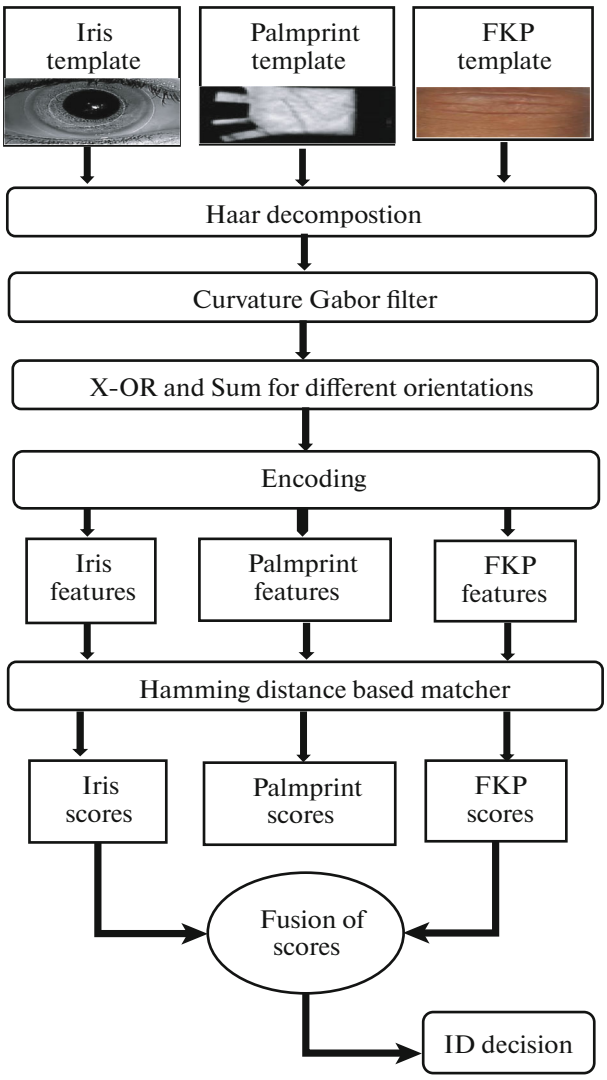


Fig. 1. The schematic block diagram for the proposed multimodal biometric system.

dimension 80 × 100 pixels, is extracted which is further normalized to 50 × 100 pixels [54].

3.2. Feature Extraction of Traits

3.2.1. Decomposition. To diminish the dimension of preprocessed images conserving the latent textural information present in them, Haar wavelet is applied, which decomposes the images into four subsections: H-H, H-L, L-H, and L-L, where L-L subsection contains greatest energy and key information of the image. So, at next level only this subsection is further processed. In this work Haar wavelet is applied at two levels, which results in downsizing the size of standardized trait by a factor of 4.This decrease in dimension benefits in augmenting the precision and speediness of the entire recognition framework.

3.2.2. Filtering. Gabor filter has the capability of human visual perception, and it is used effectively to apprehend textural information existent in the biometric images. But there are common challenges to traditional Gabor filter such as variations in illumination, enormous reflections, blurred and off-focus, which affects the performance of traditional Gabor filter. In this paper, curvature Gabor filter (CGF) is used to extract textural information existing in all the three traits, as it is proficient in analyzing textural evidences prevailing in an image. In CGF there is an extra curvature parameter as compared to traditional Gabor filter, which makes it more effective in extracting vital information from any image. Equation (1) epitomizes mathematical manifestation of the CGF [6, 59]:

$$\xi(M,\ N,\ \sigma,\ v,\ \varphi,\ c) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{\frac{m^2+n^2}{2\sigma^2}\right\} \times \exp\left\{2\pi i\left(v\left(m\cos(\phi)+n\sin(\phi)\right)+c\sqrt{m^2+n^2}\right)\right\}, \quad (1)$$

where $(M \times N)$, σ, v, φ, c symbolizes, the kernel size, standard deviation, frequency of sinusoidal function, phase offset, and curvature parameter, respectively. Figure 3 reveals the sample images of CGF filtered images of all the traits.

3.2.3. X-OR Sum coding for every orientation. In this step real and imaginary part of CGF is convolved with the image, which results in real and imaginary filtered images. Then, binarization of filtered images is done using 0 as threshold. Then, to merge the binarized real and imaginary filtered images, X-OR operator is used for all the orientations. Furthermore, a texture feature vector is obtained, by adding the outputs of X-OR operator computed for different orientations [51, 58]. The yields of X-OR and Sum operators are presented in Fig. 4.

3.2.4. Encoding. The encoding of the extracted textural feature vector is done as per following Eq. (2).



Fig. 2. Sample images of (a) normalized and segmented iris, (b) extracted palmprint ROI, and (c) extracted FKP ROI.



Fig. 3. Sample images of (a) real and imaginary parts of binarized filter response of iris, (b) real and imaginary parts of binarized filter response of palmprint, and (c) real and imaginary parts of binarized filter response of FKP.

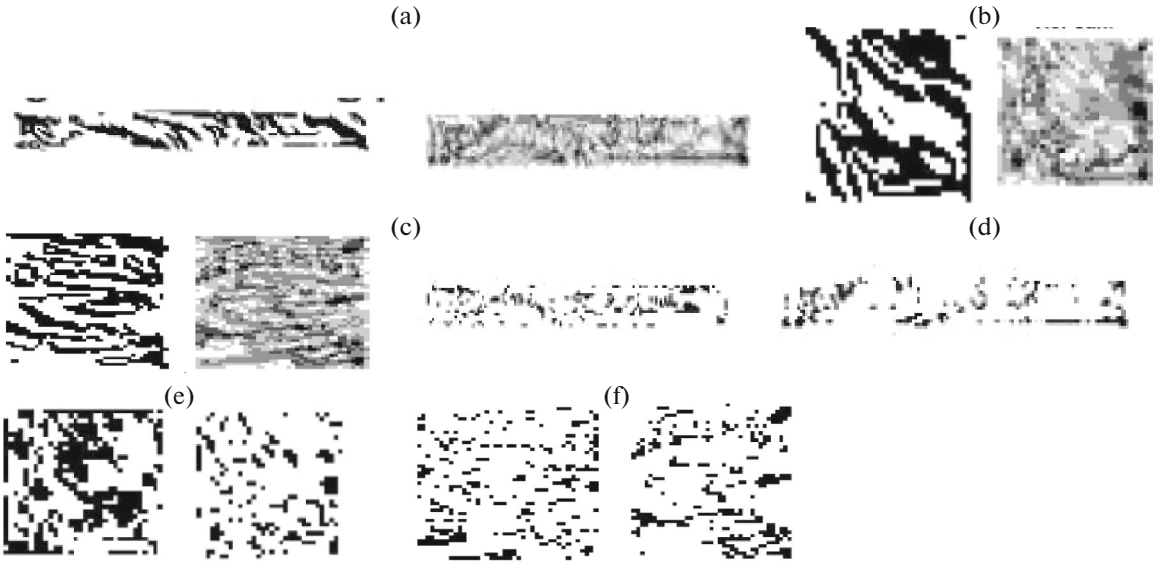


Fig. 4. Example images of (a) X-OR and X-OR Sum of iris, (b) X-OR and X-OR Sum of palmprint, (c) X-OR and X-OR Sum of FKP, (d) Bit1 and Bit2 of X-OR Sum of iris, (e) Bit1 and Bit2 of X-OR Sum of palmprint, (f) Bit1 and Bit2 of X-OR Sum of FKP.

Here, S and b epitomize the textural feature vector and number of bits essential for encoding all elements of IXSC, respectively [51, 58]

$$\mathrm{IXSC}(b) = \begin{pmatrix} 1, & \text{if } b \le S < b + \left(\dfrac{N+1}{2}\right) \\ 0, & \text{otherwise} \end{pmatrix}. \qquad (2)$$

3.3. Recognition of Traits

In this paper a Hamming distance based approach is used for matching and recognition of traits. Hamming distance is the utmost expedient approach to match feature vectors of binary images. It provides genuine and imposter scores after matching similar and non-similar images, respectively. Mathematically it can be represented as follows [51, 58]:

$$\mathrm{HD}(E,F) = \frac{\sum_{W}\sum_{L}\sum_{i=1}^{b} E_i \oplus F_i}{b \times M \times N}. \qquad (3)$$

Here, E and F are feature vectors of size $M \times N$.

3.4. Fusion Process

The efficacy of any multimodal biometric recognition framework is governed by the effectiveness of the applied fusion approach. There are various fusion schemes such as sensor level fusion, feature level fusion, score level fusion, rank level fusion and decision level fusion. In this work we have chosen score level fusion of traits as score level fusion is a realistically flexible, simple, and efficient scheme, the only prerequisite for the score level fusion is that the no. of scores should be equal for all traits to be fused. The score level fusion can be achieved using various methods like min strategy, max strategy, sum strategy, product strategy, or weighted sum strategy, provided diverse classifiers or traits. In this paper to integrate the matched scores of iris, palmprint, and finger-knuckle-print (FKP), sum and product strategies have been employed. The normalized score is designated as B_i^m for m classifiers and i volunteers. The integrated score (F_i) for user i is epitomized as follows:

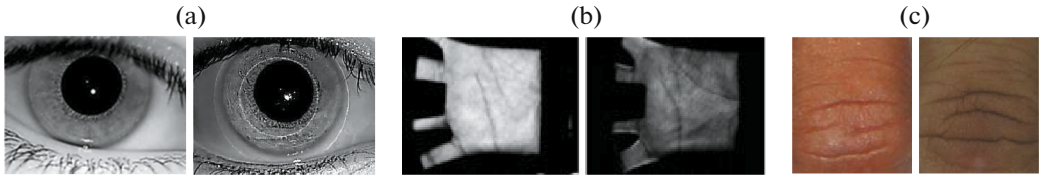


Fig. 5. Sample images from (a) IIT Delhi v1.0 Iris database, (b) PolyU-II Palmprint database, and (c) PolyUKnuckle v1 database.

- SUM fusion strategy:

$$F_i = \sum_{m=1}^{M} b_i^m, \quad \forall i; \tag{4}$$

- MAX fusion strategy:

$$F_i = \max\left(b_i^1, b_i^2, \ldots, b_i^M\right), \quad \forall i; \tag{5}$$

- MIN fusion strategy:

$$F_i = \min\left(b_i^1, b_i^2, \ldots, b_i^M\right), \quad \forall i; \tag{6}$$

- weighted SUM fusion strategy:

$$F_i = w_1 b_i^1 + w_2 b_i^2 + w_3 b_i^3 + \ldots + w_i b_i^M; \tag{7}$$

- PROD fusion strategy:

$$F_i = \sqrt{b_i^1 \times b_i^2 \times b_i^3 \times \ldots b_i^M}. \tag{8}$$

4. EXPERIMENTAL EVALUATIONS

4.1. Databases Used

In this work, for iris images, IIT Delhi v1.0 Iris Database [30] has been used which comprises 2240 images in bitmap format acquired from 224 persons i.e., 10 images from every individual with left and right eye images of size 320 × 240 pixels. For palmprint images, prevalently used PolyU-II Palmprint Database [29] has been used which consists of 7752 images acquired using CCD camera in two different sessions with diverse lighting environments from 386 subjects. Furthermore, for FKP images, contactless PolyU-Knuckle v1 database [28] has been used, which comprises 2515 finger dorsal images in bitmap format acquired from 503 subjects in two different sessions. The sample images from these three databases have been presented in Fig. 5.

4.2. Metrics for Performance Evaluation

The efficacy of the proposed approach is analysed using various metrics such as FAR (false acceptance rate), FRR (false rejection rate), EER (equal error rate), GAR (genuine acceptance rate), and DI (decidability index). FAR (false acceptance rate) can be defined as the measure of likelihood that a system mistakenly permits an illicit person to access, whom it should have barred while FAR (false rejection rate) arises when a system mistakenly rejects a genuine person to access, whom it should have allowed. EER (equal error rate) designates the total precision. It is also known as CER (crossover error rate), which can be defined as a point where the values of FRR and FAR are equivalent. The separation amongst imposter and genuine class is measured in terms of DI (decidability index), which can be described as the ratio of magnitude of difference among means of genuine (μ_g) and imposter classes (μ_i) to the root mean squares of standard deviations of imposter and genuine score which are designated as (σ_i) and (σ_g), respectively. Whereas, truly accepted users at a particular threshold is named as GAR (genuine acceptance rate). The mathematical formulas for these metrics are given as follows:

$$\mathrm{FAR} = \frac{\text{falsely_accepted}}{\text{length of imposter score}}, \tag{9}$$

$$\mathrm{FRR} = \frac{\text{falsely_rejected}}{\text{length of genuine score}}, \tag{10}$$

$$\mathrm{EER} = \frac{\mathrm{FAR} + \mathrm{FRR}}{2}, \tag{11}$$

$$\mathrm{GAR} = 1 - \mathrm{FRR}, \tag{12}$$

$$\mathrm{DI} = \frac{|\mu_g - \mu_i|}{\sqrt{\sigma_g^2 + \sigma_i^2}}. \tag{13}$$

4.3. Experimentation on Unimodal Biometric Framework

4.3.1. Iris. To gain insight into the performance of the proposed approach for an iris recognition framework, we have conducted rigorous experimentations on IIT Delhi v1.0 Iris Database. Before experimentation, every single eye image of this database has been pre-processed i.e., segmented and normalized; as a result all eye images have been transformed into a rectangular strip of 64 × 512 sizes. Table1 exhibits the performance of the iris framework for different values for curvature parameter. From the table, it can be concluded that the optimum value of curvature parameter is 0.070, as at this value the framework gives greatest DI, AUC, and GAR with least EER.

Table 1. Performance of iris recognition framework

Curvature parameter, c	DI	AUC	EER, %	GAR, %
0.030	2.6730	0.9937	2.26	96.94
0.050	2.8014	0.9949	2.08	97.35
0.070	**2.8548**	**0.9976**	**2.01**	**97.65**
0.090	2.7753	0.9975	2.33	96.97
0.110	2.7487	0.9948	2.15	96.87

Table 2. Performance of palmprint recognition framework

Curvature parameter, c	DI	AUC	EER, %	GAR, %
0.030	2.8742	0.9941	2.07	97.19
0.050	3.0023	0.9956	1.94	97.57
0.070	**3.0584**	**0.9983**	**1.86**	**97.88**
0.090	2.9747	0.9981	2.19	96.99
0.110	2.6528	0.9975	2.02	96.93

Table 3. Performance of FKP recognition framework

Curvature parameter, c	DI	AUC	EER, %	GAR, %
0.030	2.7736	0.9924	2.16	97.06
0.050	2.9018	0.9952	2.01	97.46
0.070	**2.9566**	**0.9979**	**1.93**	**97.76**
0.090	2.8750	0.9978	2.26	96.98
0.110	2.7007	0.9971	2.09	96.90

Table 4. Comparison of multimodal biometric recognition frameworks with different feature extraction techniques

Technique used	DI	AUC	EER, %	GAR, %
HOG	3.6782	0.9902	1.51	98.01
LBP	2.9823	0.9682	2.28	95.06
Log Gabor	5.8326	0.9898	1.05	98.43
Proposed approach	**6.1020**	**0.9999**	**0.81**	**99.70**

4.3.2. Palmprint. For the palmprint recognition framework, the every palmprint image from prevalently used PolyU-II Palmprint Database has been preprocessed and transformed into a square shaped strip. For extracting ROI of palmprint characterizing some reference indicators is important to depict the comparative interpretation, rotation, and scale amongst various palmprint images. In this work, Ring-small finger and index-middle finger are defined as reference milestones. The y-axis of the ROI is defined as the line fragment between two reference points, and then a square-shaped ROI is extracted using the lateral length, equal to 3/2 of the vertical line of ROI. Normalization of the extracted ROI is done using histogram equalization, which results in resized square shaped strip of dimension 144 × 144 pixels. Table 2 exhibits the performance of palmprint recognition framework for experimentations performed with different values of curvature parameter.

4.3.3. Finger-knuckle-print. Each FKP image form PoyUKnuckle v1 database has been preprocessed and for ROI extraction, the canny edge detector is employed, which works on the basis of intensity of pixels. For ROI extraction the density of pixels with high intensity is used. Therefore, a region that consists of typically edge elements and finger symmetry line of

Table 5. Performance of proposed multimodal recognition framework

Curvature parameter, c	Fusion using PROD strategy				Fusion using SUM strategy			
	DI	AUC	EER, %	GAR, %	DI	AUC	EER, %	GAR, %
0.030	5.82612	0.9993	1.01	99.47	5.4261	0.9965	1.21	98.97
0.050	5.9925	0.9994	0.88	99.64	5.5225	0.9977	1.08	99.14
0.070	**6.1020**	**0.9999**	**0.56**	**99.70**	**5.6020**	**0.9989**	**0.81**	**99.20**
0.090	5.9388	0.9996	0.90	99.41	5.4688	0.9981	1.1	98.91
0.110	5.5903	0.9997	0.83	99.56	5.1903	0.9984	1.03	99.06

dimension 80 × 100 pixels, is extracted which is further normalized to 50 × 100 pixels. Table 3 exhibits the performance of the FKP recognition framework.

4.3.4. Proposed multimodal biometric recognition framework. In the proposed multimodal biometric framework, the features of the traits are extracted using improved X-OR sum code (IXSC) in which curvature Gabor filter (CGF) is employed for extraction of the distinctive textural info existent in the traits and hamming distance based matcher for classification. The scores acquired from three traits are fused at score

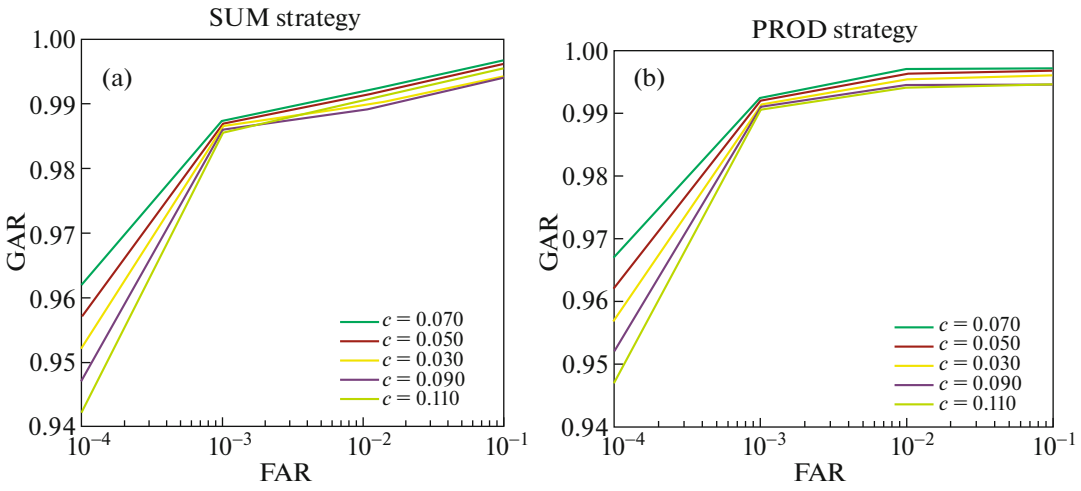


Fig. 6. ROC curves for (a) proposed approach using SUM strategy and (b) proposed approach using PROD strategy.

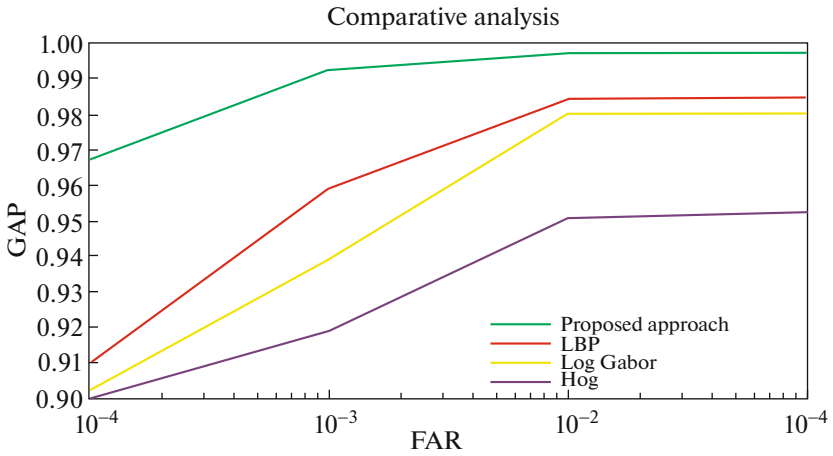


Fig. 7. ROC curves for multimodal biometric recognition frameworks with different feature extraction techniques.

level, for this product and sum strategies are employed. As Table 5 and Fig. 6 reveal, the product strategy outperforms the sum strategy with augmented performance.

4.3.5. Comparative evaluations. To validate performance of the purposed multimodal framework, a comparative analysis has been done with few eminent techniques such as LBP (local binary patterns), HOG (histogram of oriented gradients), and Log Gabor. Rigorous experimentation has been performed to develop four multimodal biometric systems using different feature descriptors. Table 4 and Fig. 7 reveal the dominance of the proposed work with greatest DI, AUC, GAR, and least EER.

5. CONCLUSIONS

This paper targets on proposing a multimodal biometric system, which can combat the obvious shortcomings of the unimodal systems. The paper utilizes three potential biometric traits, iris, palmprint and finger-knuckle-print, and presents their fusion through various score-level fusion schemes. These matching scores are attained through binary features produced using a novel curvature Gabor filter. The curvature parameter of the Gabor filter is vital in capturing the varying textural characteristics of the input traits. The comparative results, presented in the form of popular performance metrics and ROC curves, indicate that the proposed tactic outperforms the other benchmark approaches by a considerable margin, especially the low EER value of the proposed approach (0.56%) points towards its ability to be deployed in real-time and highly demanding security systems. In future, we aim to explore the deep learning convolutional neural network (CNN) based approaches to further improve the performance of the system.

COMPLIANCE WITH ETHICAL STANDARDS

This article is a completely original work of its authors; it has not been published before and will not be sent to other publications until the *PRIA* Editorial Board decides not to accept it for publication.

Conflict of Interest

The authors declare that they have no conflicts of interest.

REFERENCES

1. H. Abderrahmane, G. Noubeil, Z. Lahcene, Z. Akhtar, and D. Dasgupta, "Weighted quasi-arithmetic mean based score level fusion for-biometric systems," IET Biom. **9**, 91–99 (2020). https://doi.org/10.1049/iet-bmt.2018.5265
2. K. Aizi and M. Ouslim, "Score level fusion in multi-biometric identification based on zones of interest," J. King Saud Univ. Comput. Inf. Sci. **34**, 1498–1509 (2022). https://doi.org/10.1016/j.jksuci.2019.09.003
3. A. Alsubari, P. Lonkhande, and R. J. Ramteke, "Fuzzy-based classification for fusion of palmprint and iris biometric traits," in *Recent Trends in Signal and Image Processing*, Ed. by S. Bhattacharyya, S. Pal, I. Pan, and A. Das, Advances in Intelligent Systems and Computing, Vol. 922 (Springer, Singapore, 2019). https://doi.org/10.1007/978-981-13-6783-0_11
4. B. Ammour, T. Bouden, and L. Boubchir, "Face–iris multi-modal biometric system using multi-resolution Log-Gabor filter with spectral regression kernel discriminant analysis," IET Biom. **7**, 482–489 (2018). https://doi.org/10.1049/iet-bmt.2017.0251
5. B. Attallah, A. Serir, Y. Chahir, and A. Boudjelal, "Application of BSIF, Log-Gabor and mRMR transforms for iris and palmprint based Bi-modal identification system," in *Int. Conf. on Advanced Technolies for Signal and Image Processing (ATSIP), Fez, Morocco, 2017* (IEEE, 2017), pp. 1–6. https://doi.org/10.1109/ATSIP.2017.8075600
6. N. Bala, R. Vyas, R. Gupta, and A. Kumar, "Iris recognition using improved xor-sum code," in *Security and Privacy*, Ed. by P. Stănică, S. Gangopadhyay, and S. K. Debnath, Lecture Notes in Electrical Engineering, Vol. 744 (Springer, 2021), pp. 107–117. https://doi.org/10.1007/978-981-33-6781-4_9
7. I. Boucherit, M. O. Zmirli, H. Hentabli, and B. A. Rosdi, "Finger vein identification using deeply-fused Convolutional Neural Network," J. King Saud Univ. Comput. Inf. Sci. **34**, 646–656 (2022). https://doi.org/10.1016/j.jksuci.2020.04.002
8. L. Boussaad and A. Boucetta, "An effective component-based age-invariant face recognition using Discriminant Correlation Analysis," J. King Saud Univ. Comput. Inf. Sci. **34**, 1739–1747 (2020). https://doi.org/10.1016/j.jksuci.2020.08.009
9. K. W. Bowyer, K. P. Hollingsworth, and P. J. Flynn, "A survey of iris biometrics research: 2008–2010," in *Handbook of Iris Recognition*, Ed. by M. Burge and K. Bowyer, Advances in Computer Vision and Pattern Recognition (Springer, London, 2012), pp. 15–54. https://doi.org/10.1007/978-1-4471-4402-1_2
10. F. Cherifi, K. Amroun, and M. Omar, "Robust multimodal biometric authentication on IoT device through ear shape and arm gesture," Multimedia Tools Appl. **80**, 14807–14827 (2021). https://doi.org/10.1007/s11042-021-10524-9
11. J. Daugman, "High confidence visual recognition of persons by a test of statistical significance," IEEE Trans. Pattern Anal. Mach. Intell. **15**, 1148–1161 (1993). https://doi.org/10.1109/34.244676
12. J. Daugman, "How iris recognition works," in *The Essential Guide to Image Processing*, Ed. by A. I. Bovik, 2nd ed. (Academic Press, 2009), pp. 715–739. https://doi.org/10.1016/B978-0-12-374457-9.00025-1
13. J. Daugman and C. Downing, "Radial correlations in iris patterns, and mutual information within IrisCodes," IET Biom. **8**, 185–189 (2019). https://doi.org/10.1049/iet-bmt.2018.5199

14. D. V. R. Devi and K. N. Rao, "A multimodal biometric system using partition based DWT and rank level fusion," in *IEEE Int. Conf. on Computational Intelligence and Computing Research (ICCIC), Chennai, India, 2016* (IEEE, 2016), pp. 1–5. https://doi.org/10.1109/ICCIC.2016.7919543
15. J. Grover and M. Hanmandlu, "Personal identification using the rank level fusion of finger-knuckle-prints," Pattern Recognit. Image Anal. **27**, 82–93 (2017). https://doi.org/10.1134/S1054661817010059
16. K. Gupta, G. Singh, and W. Kapil, "Novel approach for multimodal feature fusion to generate cancelable biometric," Visual Comput. **37**, 1401–1413 (2020). https://doi.org/10.1007/s00371-020-01873-x
17. R. Gupta, M. Khari, D. Gupta, and R. G. Crespo, "Fingerprint image enhancement and reconstruction using the orientation and phase reconstruction," Inf. Sci. **530**, pp. 201–218, 2020. https://doi.org/10.1016/j.ins.2020.01.031
18. S. Hariprasath and M. Santhi, "Bimodal biometric pattern recognition system based on fusion of iris and palmprint using multi-resolution approach," Signal, Image Video Process. **14**, 519–527 (2020). https://doi.org/10.1007/s11760-019-01573-7
19. K. P. Hollingsworth, K. W. Bowyer, and P. J. Flynn, "Improved iris recognition through fusion of hamming distance and fragile bit distance," IEEE Trans. Pattern Anal. Mach. Intell. **33**, 2465–2476 (2011). https://doi.org/10.1109/TPAMI.2011.89
20. A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities," Pattern Recognit. Lett. **79**, 80–105 (2016). https://doi.org/10.1016/j.patrec.2015.12.013
21. A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," IEEE Trans. Circuits Syst. Video Technol. **14**, 4–20 (2004). https://doi.org/10.1109/TCSVT.2003.818349
22. D. K. Jain, X. Lan, and R. Manikandan, "Fusion of iris and sclera using phase intensive rubbersheet mutual exclusion for periocular recognition," Image Visual Comput. **103**, 104024 (2020). https://doi.org/10.1016/j.imavis.2020.104024
23. G. Jaswal and R. C. Poonia, "Selection of optimized features for fusion of palm print and finger knuckle-based person authentication," Exp. Syst. **38**, e12523 (2020). https://doi.org/10.1111/exsy.12523
24. Z. Jiang, C. Guo, J. Zang, G. Lu, and D. Zhang, "Features fusion of multichannel wrist pulse signal based on KL-MGDCCA and decision level combination," Biomed. Signal Process. Control **57**, 101751 (2020). https://doi.org/10.1016/j.bspc.2019.101751
25. T. Joseph, S. A. Kalaiselvan, S. U. Aswathy, R. Radhakrishnan, and A. R. Shamna, "A multimodal biometric authentication scheme based on feature fusion for improving security in cloud environment 1 Introduction," J. Ambient Intell. Humanized Comput. **12**, 6141–6149 (2020). https://doi.org/10.1007/s12652-020-02184-8
26. D. Karmakar, P. Mukherjee, and M. Datta, "Spoofed facial presentation attack detection by multivariate gradient descriptor in micro-expression region," Pattern Recognit. Image Anal. **31**, 285–294 (2021). https://doi.org/10.1134/S1054661821020097
27. W. K. Kong and D. Zhang, "Palmprint texture analysis based on low-resolution images for personal authentication," in *Int. Conf. on Pattern Recognition, Quebec City, Canada, 2002* (IEEE, 2002), pp. 807–810. https://doi.org/10.1109/ICPR.2002.1048142
28. A. Kumar, "Importance of being unique from finger dorsal patterns: Exploring minor finger knuckle patterns in verifying human identities," IEEE Trans. Inf. Forensics Secur. **9**, 1288–1298 (2014). https://doi.org/10.1109/TIFS.2014.2328869
29. A. Kumar, "Toward more accurate matching of contactless palmprint images under less constrained environments," IEEE Trans. Inf. Forensics Secur. **14**, 34–47 (2019). https://doi.org/10.1109/TIFS.2018.2837669
30. A. Kumar and A. Passi, "Comparison and combination of iris matchers for reliable personal authentication," Pattern Recognit. **43**, 1016–1026 (2010). https://doi.org/10.1016/j.patcog.2009.08.016
31. A. M. Mayya and M. Saii, "Iris and palmprint decision fusion to enhance human recognition," Int. J. Comput. Sci. Trends Technol. **5** (5), 42–46 (2017).
32. H. Mehrotra, P. K. Sa, and B. Majhi, "Fast segmentation and adaptive SURF descriptor for iris recognition," Math. Comput. Modell. **58** (1–2), 132–146 (2013). https://doi.org/10.1016/j.mcm.2012.06.034
33. M. Sharma, R. Baraskar, and S. Agrawal, "A comparative analysis of unimodal and multimodal biometric systems," Int. J. Adv. Res. Comput. Sci. **8** (5), 297–306 (2017).
34. H. Naderi, B. H. Soleimani, S. Matwin, B. N. Araabi, and H. Soltanian-Zadeh, "Fusing iris, palmprint, and fingerprint in a multi-biometric recognition system," in *13th Conf. Computer Robot Vision (CRV), Victoria, Canada, 2016* (IEEE, 2016), pp. 327–334. https://doi.org/10.1109/CRV.2016.18
35. H. Naderi, B. H. Soleimani, and S. Matwin, "Manifold learning of overcomplete feature spaces in a multimodal biometric recognition system of iris and palmprint," in *14th Conf. Computer Robot Vision (CRV), Edmonton, Canada, 2017* (IEEE, 2017), pp. 191–196. https://doi.org/10.1109/CRV.2017.29
36. S. K. Naji, "Performance of decision fusion based human identification system," **17** (9), 212–227 (2020).
37. L. Nanni and A. Lumini, "A multi-matcher system based on knuckle-based features," Neural Comput. Appl. **18**, 87–91 (2009). https://doi.org/10.1007/s00521-007-0160-4
38. P. Punyani, R. Gupta, and A. Kumar, "Human age-estimation system based on double- level feature fusion of face and gait images," Int. J. Image Data Fusion **9**, 222–236 (2018). https://doi.org/10.1080/19479832.2018.1423644
39. P. Punyani, R. Gupta, and A. Kumar, "Neural networks for facial age estimation: A survey on recent advances," Artif. Intell. Rev. **53**, 3299–3347 (2020). https://doi.org/10.1007/s10462-019-09765-w

40. H. Purohit and P. K. Ajmera, "Optimal feature level fusion for secured human authentication in multimodal biometric system," Mach. Vision Appl. **32**, 24 (2021). https://doi.org/10.1007/s00138-020-01146-6

41. S. A. El Rahman, "Multimodal biometric systems based on different fusion levels of ECG and fingerprint using different classifiers," Soft Comput. **24**, 12599–12632 (2020). https://doi.org/10.1007/s00500-020-04700-6

42. G. Rajagopal and R. Palaniswamy, "Performance evaluation of multimodal multifeature authentication system using KNN classification," Sci. World J. **2015**, 762341 (2015). https://doi.org/10.1155/2015/762341

43. C. Ramachandran and D. Sankar, "Score level based fusion method for multimodal biometric recognition using palmprint and iris," in *Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), Cochin, India, 2020* (IEEE, 2020), pp. 281–286. https://doi.org/10.1109/ACCTHPA49271.2020.9213216

44. M. E. Rane and U. S. Bhadade, "Multimodal score level fusion for recognition using face and palmprint," Int. J. Electr. Eng. Educ. (2020). https://doi.org/10.1177/0020720920929662

45. P. P. Sarangi, D. R. Nayak, M. Panda, and B. Majhi, "A feature-level fusion based improved multimodal biometric recognition system using ear and profile face," J. Ambient Intell. Humanized Comput. **13**, 1867–1898 (2022). https://doi.org/10.1007/s12652-021-02952-0

46. M. Singh, R. Singh, and A. Ross, "A comprehensive overview of biometric fusion," Inf. Fusion **52**, 187–205 (2019). https://doi.org/10.1016/j.inffus.2018.12.003

47. R. Singh, A. Ross, and K. W. Bowyer, "Special issue on information fusion in biometrics," Inf. Fusion **32** (PB), 1–2 (2016). https://doi.org/10.1016/j.inffus.2016.04.004

48. M. Sivarathinabala, S. Abirami, M. Deivamani, and M. Sudharsan, "A smart security system using multimodal features from videos," Pattern Recognit. Image Anal. **29**, 89–98 (2019). https://doi.org/10.1134/S1054661819010218

49. S. Soviany, "A biometric identification system with kernel SVM and feature-level fusion," in *12th Int. Conf. on Electronics, Computers, and Artificial Intelligence (ECAI), Bucharest, 2020* (IEEE, 2020), pp. 1–6. https://doi.org/10.1109/ECAI50035.2020.9223188

50. R. Srivastava, "A score level fusion approach for multimodal biometric fusion," Int. J. Sci. Technol. Res. **9**, 4241–4245 (2020).

51. D. Tamrakar and P. Khanna, "Palmprint verification with XOR-SUM Code," Signal, Image Video Process. **9**, 535–542 (2015). https://doi.org/10.1007/s11760-013-0475-9

52. S. D. Thepade and R. K. Bhondave, "Bimodal biometric identification with Palmprint and Iris traits using fractional coefficients of Walsh, Haar, and Kekre transforms," in *Int. Conf. Communication, Information & Computing Technology (ICCICT), Mumbai, India, 2015* (IEEE, 2015), pp. 15–18. https://doi.org/10.1109/ICCICT.2015.7045729

53. R. B. Trabelsi, A. D. Masmoudi, and D. Sellami Masmoudi, "A real time of an automatic finger vein recognition system," Pattern Recognit. Image Anal. **28**, 430–438 (2018). https://doi.org/10.1134/S1054661818030173

54. K. Usha and M. Ezhilarasan, "Personal recognition using finger knuckle shape oriented features and texture analysis," J. King Saud Univ. Comput. Inf. Sci. **28**, 416–431 (2016). https://doi.org/10.1016/j.jksuci.2015.02.004

55. V. S. Amritha and J. Aravinth, "Matcher performance-based score level fusion schemes for multi-modal biometric authentication system," in *6th Int. Conf. on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020* (IEEE, 2020), pp. 79–85. https://doi.org/10.1109/ICACCS48705.2020.9074446

56. R. Vinothkanna and A. Wahi, "A multimodal biometric approach for the recognition of finger print, palm print and hand vein using fuzzy vault Amitabh Wahi," Int. J. Biomed. Eng. Technol. **33** (1), 54–76 (2020).

57. R. Vyas, T. Kanumuri, G. Sheoran, and P. Dubey, "Efficient iris recognition through curvelet transform and polynomial fitting," Optik **185**, 859–867 (2019). https://doi.org/10.1016/j.ijleo.2019.04.015

58. R. Vyas, T. Kanumuri, and G. Sheoran, "Iris recognition using 2-D Gabor filter and XOR-SUM code," in *1st India Int. Conf. on Information Processing (IICIP), Delhi, 2016* (IEEE, 2016), pp. 1–5. https://doi.org/10.1109/IICIP.2016.7975369

59. H. Wang, M. Du, J. Zhou, and L. Tao, "Weber local descriptors with variable curvature Gabor filter for finger vein recognition," IEEE Access **7**, 108261–108277 (2019). https://doi.org/10.1109/ACCESS.2019.2928472

60. Q. Xiong, X. Zhang, X. Xu, and S. He, "A modified chaotic binary particle swarm optimization scheme and its application in face-iris multimodal biometric identification," Electronics **10**, 1–17 (2021). https://doi.org/10.3390/electronics10020217

61. X. Zhang, D. Cheng, P. Jia, Y. Dai, and X. Xu, "An efficient android-based multimodal biometric authentication system with face and voice," IEEE Access **8**, 102757–102772 (2020). https://doi.org/10.1109/ACCESS.2020.2999115



Neeru Bala is Research Scholar in Department of Electronics and Communication Engineering at Amity University Haryana. She completed her MTech in Electronics and Communiaction Engineering from Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonepat, Haryana, India in 2009. She has more than 10 publications in national/international conferences/journals, attended/organized 20 faculty development workshops held time to time. Her primary research interests are signal and image processing, and biometrics.



Rashmi Gupta is presently working as Professor in Electronics and Communication Engineering Department in Netaji Subhas Univerity of Technology, East Campus (Formerly AIACTR) Goverment of NCT of Delhi. She is also holding position of Founder Director of Incubation Center AIACTR-IRF funded by Delhi Govt. She is Enterpreneurship Ambassador of Institution's Innovation Council Initiative by MHRD. Prior to this she has 8 yr of industry experience and 20 yr of teaching experience in various academic institutes. She has done her ME and PhD degree in Electronics and Communication Engineering from Delhi College of Engineering, Delhi University.

Gupta has authored over 90 research papers in various renowned international journal and conferences. She has also published five patents. Prof. Gupta is associated with many International research organizations as editorial board member and reviewer. She is Fellow Member of IETE and Senior Member of IEEE. Her primary research interests are signal and image processing, computer vision, artificial intelligence, wireless communication, IoT, and security.



Anil Kumar is faculty in Department of Electronics and Communication Engineering at Amity University Haryana. He completed his MTech in Instrumentation engineering from Panjab University Chandigarh in 2008. He did his PhD from Amity University Haryana in the field of nanotechnology. Before joining Amity University (September 2010), he served Gurgaon College of Engineering, Bilaspur (HR) from July 2008–Sept. 2010 as Assistant Professor. He is member of various technical bodies including IETE, IACSIT, IAENG, IAENG Society of Electrical Engineering, IAENG Society of Imaging Engineering and SIE. He has more than 55 publications in national/international conferences/journals, attended/organized 40 faculty development workshops held time to time. During his career, he has taken number of administrative tasks including organizing, mentoring various events, coordinating various courses, etc.

2022 2nd Asian Conference on Innovation in Technology (ASIANCON)
Pune, India. Aug 26-28, 2022

An Inner Round Pipeline Architecture Hardware Core for AES

Archit Jain
Dept. of Electronics and Communication
Delhi Technological University
Delhi, India
archit.jain38@gmail.com

Divyanshu Jain
Dept. of Electronics and Communication
Delhi Technological University
Delhi, India
2000divyanshujain@gmail.com

Arpan Katiyar
Dept. of Electronics and Communication
Delhi Technological University
Delhi, India
arpankatiyardtu@gmail.com

Prof. Gurjit Kaur
Dept. of Electronics and Communication
Delhi Technological University
Delhi, India
gurjitkaur@dtu.ac.in

2022 2nd Asian Conference on Innovation in Technology (ASIANCON) | 978-1-6654-6851-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ASIANCON55314.2022.9909114

***Abstract*—The following article presents an inner round architecture for the AES Encryption Scheme suitable for implementation on FPGAs and as ASICs. The uniformity between the encryption and decryption hardware makes them suitable for implementation as separate or co-existing blocks as required. The modular approach of our architecture allows for different encryption/decryption core configurations providing a compact, scalable implementation that is suitable for applications that may demand compact yet high performant hardware. The architecture employs a combinational S-Box forming a crucial step in the parallel operation of the hardware. For an operating frequency of 278.5 MHz, the hardware achieves a high throughput of about 3.5 gigabits per second (GBps).**

***Index Terms*—Encryption, AES, Cryptography, Hardware**

I. Introduction

Encryption has become an absolutely indelible component of privacy and security in modern digital communication. The Advanced Encryption Standard (AES) (originally the Rijndael Algorithm) [1] [2] defined by Daemen and Rijmen at the National Institute of Standards Technology (NIST) in 2001 when a requirement for an already prevailing encryption scheme, the Data Encryption Standard (DES) surfaced, as it was becoming more and more insecure to brute force attacks due to algorithmic and mathematical shortcomings. Since then, this has become a benchmark cryptographic Encryption standard as the most secure solutions prevailing at this time.

In order to achieve higher speeds of encryption/decryption, this standard is being constantly improved and implemented in hardware and software to encrypt sensitive data. Hardware implementation performs better in all aspects hence various ASIC and FPGA implementations for AES Algorithm are being developed [12]- [16] in order to create a better architecture for Encryption and Decryption matching to the standard speed of data transmission such that communication can work with encryption without causing a bottleneck on speeds. Apart from enhancing bits/sec of encryption various works are dedicated for low power consumption and compact hardware architecture designs for AES implementation. [16].

We propose an inner round pipeline architecture for implementation at the silicon level for the AES Algorithm, to increase the performance of Encryption by maximum utilisation of hardware at each timestep. Advanced selective pipelined architecture to provide each step-specific hardware portion occupied with at least one block of data simultaneously, equivalent to a linear pipelined approach, hence enhancing overall throughput and minimising unoccupied state power wastage. The paper is divided into five sections: Section II introduces the algorithm itself; Section III discusses our implementation, Section IV puts forth simulation results, and Section V concludes the article.

II. AES Overview

The AES Algorithms set a constant length of 128 bits for the state, input, and output blocks of data. The Nb = 4 characterises (each of a 4-byte words). The secret key could be 16, 24, 32 bytes in length. Finally, N_r indicates the number of rounds to be done on each block, with N_r equal to 10 for a 16-byte key, 12 for a 24-byte key, and 14 for a 32-byte key. As a result, the number of rounds is proportional to the size of the key.

The algorithm proceeds in rounds and each round comprises four transformations: 1) By using a fixed lookup, substitution at the byte level. 2) Using different offsets to move rows, 3) mathematically mixing data inside each column, and 4) addition of a Round Key

The Cipher operates on the state which is initialised to the input at the beginning. The algorithm initiates with a Round Key addition(bitwise xor), following which the State array is passed through N_r rounds, with the final round being differing in a single operation. The resulting state is the final output. The round key used at each bitwise xor with the round input, is expanded using the Key Expansion procedure highlighted in

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:10:44 UTC from IEEE Xplore. Restrictions apply.

the following sub sections. All these operations are invertible, and hence to obtain the inverse-cipher, the inverses of each of the operations are applied in the reversed order.

A. Sub Bytes

This is a non-linear substitution utilising two operations on each byte; The multiplicative inverse of the byte in $GF(2^8)$ (00 mapped to 00), and second using an affine transformation as specified in [1].

B. Row Shift

Each block of data is processed such that, the final three rows are moved in a cyclic fashion across various distinct offsets in this transformation [1].

C. Mix Columns

This sub-process works on the state at each column, where each column is represented by a polynomial of the fourth degree, as a function of previous column inputs [1].

D. Add Round Key

It is simply bit-wise xor of input with round specific expanded key generated by key expanding module. A key is added to the state register. Each round uses a unique key [1].

E. Inverse Sub Bytes

The inverse of the SubBytes() Transform is InvSubBytes(). This is defined similar to the original SubBytes(), i.e. by applying the inverse of affine transformation and multiplying the result by the multiplicative inverse [1].

F. Inverse Row Shift

This is simply the inverse transformation of the Shift Rows transformation. The first row is again left non-transformed as is the case with the Shift-Row operation [1].

G. Inverse Mix Columns

It's simply the inverse of the Mix Columns transformation, in which each column is represented using a four-term polynomial multiplied with a fixed polynomial $a^{-1}(x)$ and taking modulo $x^4 + 1$ [1].

H. Inverse Add Round Key

InvAddRoundKey is the same as AddRoundKey since AddRoundKey() is its own inverse.

I. Key Expansion

The Cipher Key is used to produce a sequence of keys using an algorithm called Key Routine or the Key schedule. This results in the $N_b(N_r + 1)$ key words. The keys that result comprise 4-byte words, each of which is represented by $[w_i]$. Here,

$$0 \leq i < N_b(N_r + 1)$$

The key routine proceeds with RotWord operation on the last word, followed by the SubWord operation, addition of a certain constant value also called the round constant and then addition of the previous round key (4 words). RotWord takes a word as input, applies a cyclic permutation, and outputs another word. SubWord is the S-Box operation and is used to create an output word from four input bytes. RCon[i] is made up of round constants and is denoted as $[x^{i-1}, 00, 00, 00]$, with x as (02) in the field $GF(2^8)$ and $i \leq 1$. A small deviation from the usual definition here, is that if $i - 4$ is a multiple of N_k, a SubWord() transformation is also done to the word $w[i-1]$ prior to XORing for $N_k = 8$ (256 bit key).

III. Implementation

A. Architecture

The proposed architecture scheme of the hardware is shown in figure 1. The architecture remains compact as well as performant as the implementation of the blocks itself remains compact. The sub blocks are tightly integrated with each other without sacrificing the potential to have parallel encryption/decryption blocks.

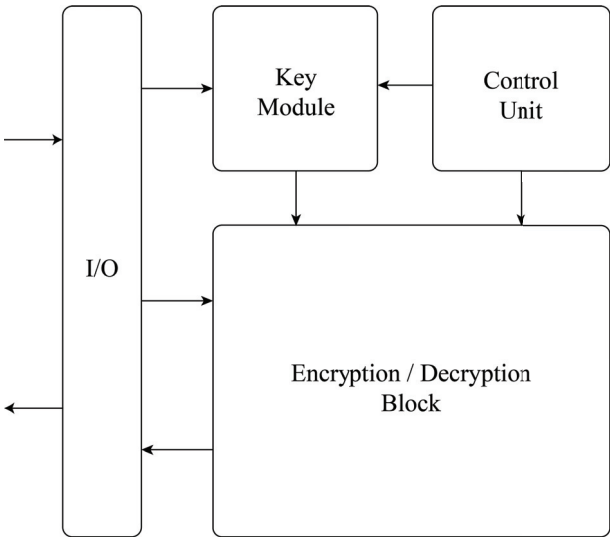


Fig. 1. Proposed Architecture

B. Key Module

The key module is responsible to generate the key schedule as specified by the standard. For the encryption, the keys were computed on the fly. The decryption process however requires the expanded keys in the reverse order. A straightforward way to implement this is to build a memory that stores all the expanded keys. This approach however requires $128*N_r$ number of registers.

The key routine specification uses a round constant to remove symmetries in the inter round keys. It can be noted that if one complete round key and its corresponding round constant is known, the complete key schedule can be determined. This allows us to store only the last key in the schedule and generate all the previous keys on the fly.

The key expansion hardware in either direction is purely combinational and uses only output registers between rounds

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:10:44 UTC from IEEE Xplore. Restrictions apply.

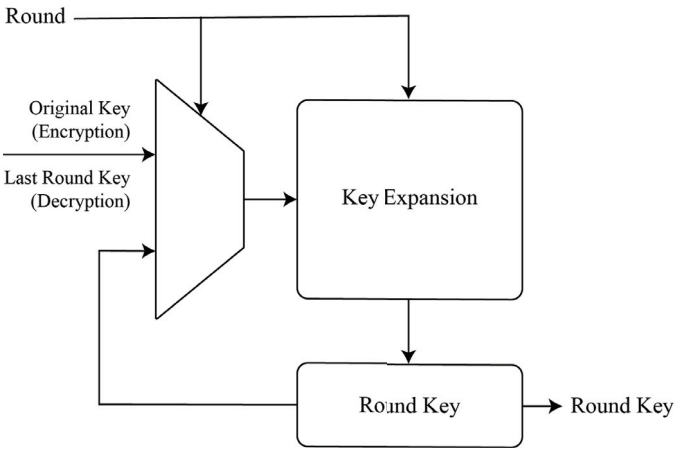


Fig. 2. Key Module

and to store the original and last round key. This implementation requires 136 XOR gates, and 4 S-Boxes in the place of 128*N_r flops (when N_r = 10, 1280 Flops).

C. Encryption/Decryption Block

The AES "round" is made up of four major operations that must be performed in a recurring order. When these rounds are implemented in combinational order, the result is a longer combinational path delay, lowering throughput. By dividing the data path into three pipelined stages and adding registers in between, the critical path delay is reduced, resulting in increased operating frequency. The architecture is shown in

As a result, by trading off on latency, the system operates at a higher clock frequency, resulting in increased overall throughput in bits encrypted per second.

Minimum time in general approach

$$T_{min} = T_{block1} + T_{block2} + T_{block3}$$

Minimum time in pipelined approach

$$T_{min} = min(T_{block1}, T_{block2}, T_{block3})$$

Maximum Operable frequency

$$F_{max} = 1/T_{min}$$

where, T_{block1} = Critical path delay for block1 i.e., SBOX subsystem; T_{block2} = Critical path delay for block2 i.e., shiftRow and MixCol subsystem; T_{block3} = Critical path delay for block3 i.e., AddRoundKey subsystem.

Add round key is a simple xor of the data stream with the expanded round key and is implemented using xor gates. The row shift operation is at its simplest, linear shift of n bytes where n is row number based on position. Such an operation is wired directly to the required position to the next state register/operation circuit. This avoids the need to implement the operation using shift registers, thereby reducing the hardware and the number of clock cycles per round. This transform requires no gate delay.

Mix column is the matrix multiplication of each row by fixed matrix as discussed in the previous section. This complex matrix operation is simplified by constructing a Boolean expression for each cell in the column, resulting in a combinational circuit for the same. Multiplication is implemented at the byte level by reducing the complex multiplication to a simple combination of shift and xor.

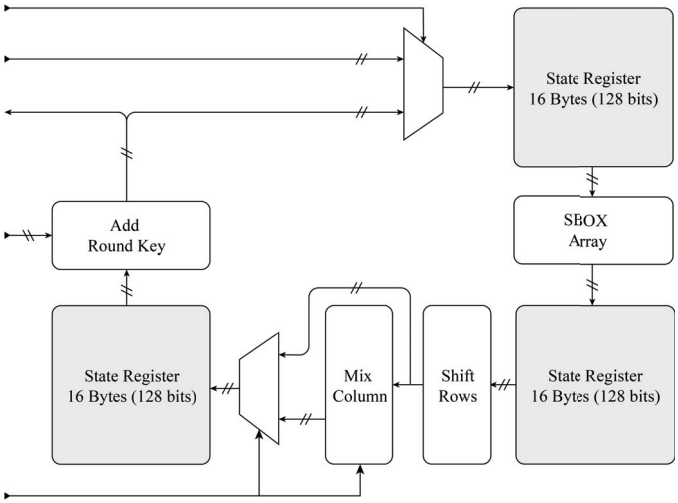


Fig. 3. Encryption/Decryption Block

The S-Box is one of the most crucial components of the algorithm. Simple implementations of the S-Box use lookup tables to read only memories, flash etc. However, our architecture requires a combinational implementation as the substitution should complete in a single cycle. Different types of combinational implementations have been proposed [3] We implemented the circuit as proposed by [17]. Here, the byte is converted to an equivalent element in the field $GF(2^4)^2$ which is inherently easier to implement in combinational hardware using exclusive-OR gates. 16 such S-boxes are used for 16 substitutions to take place at each clock cycle.

This architectural approach resulted in a maximum clock frequency of roughly 278 MHz by considering the maximum data path offered by S-Box. After analysing the remaining 3 transforms, we combined the mix column and shift rows into a single clock cycle so as to reduce the number of required clock cycles from 4 to 3 without affecting the operating clock frequency which depends on the delay of the S-Box. By combining stages, the registers used are reduced.

As shown in figure 3, the encryption data path consists of three stages and 128-bit registers are introduced in between each stage. At the end of each round 3 clock cycles are utilised as the data passes through these 3 registers and the intermediate stages.

The control unit can be altered to generate output from this semi pipelined block after 10, 12, 14 according to the key size. The inverse shift row and inverse S-box have a similar construction. The inverse mix column's matrix multiplication is reduced in the same way as encryption was reduced using combinational Boolean reduction but with different values in the matrix.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:10:44 UTC from IEEE Xplore. Restrictions apply.

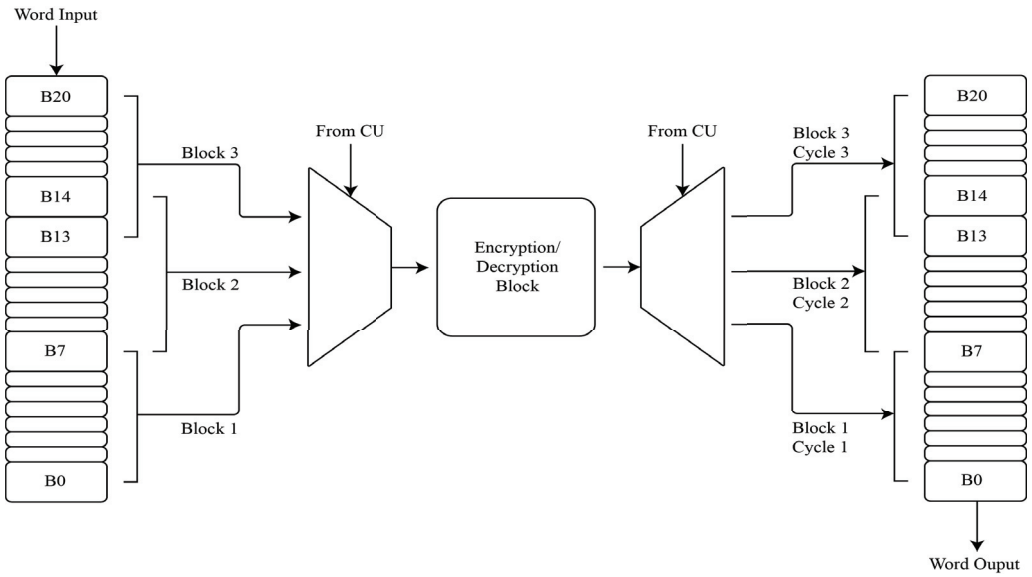


Fig. 4. I/O Pipelines

D. Input/Output Unit

The I/O unit (Figure 4) handles all the incoming and outgoing data which forms a continuous stream of 16-bit words on either side. However, the data path inside the actual encryption circuit is 128 bit (16 bytes, one data block) and pipelines are used to affect the data width change.

The key is taken as a serial input and stored inside the unit and cannot be accessed by any other block save the key module.

E. Control Unit

The control unit is responsible for control and synchronising operations between blocks. Signals from the control unit control the flow of data into the next round or out of the block to the I/O unit, entry of a new encryption key, expansion of the key using key schedule in the key module and also the 16-bit width I/O pipes. It consists of control flops, flag registers and a couple of counters for rounds and stages.

IV. Results

On an FPGA, the proposed encryption core architecture and its accompanying input/output interface were synthesised. At a clock frequency of 278.5MHz, we were able to attain a throughput of 3.5Gbps. Since the design itself is compact, about 4000 CLBs were used after inter boundary optimisations. The throughput may be further increased by using multiple encryption/decryption blocks with minimum addition to the control hardware.

It can be seen that the throughput is appreciably high with respect to the already implemented designs, however the hardware utilisation is very low thus making it suitable for use in designs with requirements for specific hardware for encryption such as a TPM, etc.

TABLE I
Hardware Utilisation for Different Designs

Design	Slice LUTs	Clock (MHz)	T/S
[3]	11022	168.4	1.96
[4]	5177 / round	168.3	4.2
[5]	16938	184.2	1.39
[6]	19125	128.07	0.67
[7]	86806	250	0.37
[8]	10762	305.1	3.66
[9]	22994	576.06	3.21
[10]	8901	500	7.19
[11]	20155	180.39	1.15
Our Design	4505	278.5	0.37

References

[1] M.R.M.Rizk, “Optimized area and optimized speed hardware implementation of AES on FPGA”, 2nd International Design and Test Workshop, 2007, pp.207-217.

[2] National Institute of Standards and Technology, Federal Information Processing Standard 197: The Advanced Encryption Standard (AES), 2001, http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf.

[3] X. Shang, K. K. Parhi, “High speed VLSI architectures for the AES algorithm”, IEEE Transaction on Very Large Scale Integration Systems, Vol.12, No.9, 2004, pp.957-967.

[4] A. Hodjat, I. Verbauwhede,” A 21.54 Gbits/s fully pipelined AES processor”, IEEE Symposium on Field Programmable Custom Computing Machines (FCCM), 2004.

[5] J.Zambreno, D.Nguyen, A.Choudhary, “Exploring Area/Delay tradeoffs in an AES FPGA implementation”, 2004, pp.575-585.

[6] Nalini.C, Nagaraj, A. Mohan, Poornaih D.V, V.D.Kulkarni, “An FPGA based performance analysis of pipelining and u$N_r$olling of AES algorithm”,Proceedings of International Conference on Advanced Computing and Communications, 2006, pp.477-482.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:10:44 UTC from IEEE Xplore. Restrictions apply.

[7] Chih-Penh Fan, Ju-Kui Hwang,"Implementations of high throughput sequential and fully pipelined processors on FPGA", Proceedings of International Symposium on Intelligent Signal Processing and Communication systems, 2007, pp.353-356.
[8] I. Hammad, K. El-sankary, E. El-Masry, "High speed AES encryptor with efficient merging techniques", IEEE Embedded systems Letters, Vol.2 Issue 3, 2010, pp.67-71.
[9] S. Qu, G. Shou, Y. Hu, Z. Guo, Z. Qian, "High throughput, pipelined implementation of AES on FPGA", International symposium on Information Engineering and Electronic Commerce, 2009, pp.542-545.
[10] D. Chen, G. Shou, Y. Hu, Z. Guo, "Efficient architecture and implementation of AES", 3rd International Conference on Advanced Computer Theory and Engineering, 2010, pp.295-298.
[11] W. Stallings, Cryptography and Network Security, 3rd ed., Upper Saddle River,N.J.: Prentice Hall, 2003.
[12] A. Satoh, S. Morioka, K. Takano, and S. Munetoh, "A compact Rijndael hardware architecture with S-box optimization," in Proc. ASIACRYPT 2001, 2001, pp. 239–254.
[13] F.K. Gurkaynak et al., "A 2 Gb/s balanced AES crypto-chip implementation," in Proc. ACM Great Lakes Symp. VLSI (GLSVLSI '04), Boston, Mass., Apr. 2004,pp. 39–44.
[14] A. Elbirt, W. Yip, B. Chetwynd, and C. Paar, "An FPGA-based performance evaluation of the AES block cipher candidate algorithm finalists," IEEE Trans. VLSI Syst.,vol. 9, Aug. 2001, pp. 545–557.
[15] P. Chodowiec, P. Khuon, and K. Gaj, "Fast implementations of secret key block ciphers using mixed inner- and outer-round pipelining," in Proc. Int. Symp. Field Programmable Gate Arrays, Feb. 2001, pp. 94–102.
[16] P. Chodowiec and K. Gaj, "Very compact FPGA implementation of the AES algorithm," in Proc. 5th Int. Workshop Cryptographic Hardware and Embedded Systems (CHES 2003), 2003, pp. 319–333.
[17] Akashi Satoh, Sumio Morioka, Kohji Takano, and Seiji Munetoh "A compact Rijndael hardware architecture with S-box optimization"

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:10:44 UTC from IEEE Xplore. Restrictions apply.

2022 2nd Asian Conference on Innovation in Technology (ASIANCON)
Pune, India. Aug 26-28, 2022

An Optimized XGBoost approach for Predicting Progression of Hepatitis C using Hyperparameter Tuning and Feature Interaction Constraint

Kaustubh Ranjan Singh
Dept. of Electronics & Comm. Engineering
Delhi Technological University(DTU)
Delhi, India
kaustubh@dtu.ac.in

Ritik Gupta
Dept. of Electronics & Comm. Engineering
Delhi Technological University(DTU)
Delhi, India
ritikgupta_2k18ec136@dtu.ac.in

Ritik Kumar Kadian
Dept. of Electronics & Comm. Engineering
Delhi Technological University(DTU)
Delhi, India
ritikkumarkadian_2k18ec137@dtu.ac.in

Ritwik Singh
Dept. of Electronics & Comm. Engineering
Delhi Technological University(DTU)
Delhi, India
ritwiksingh_2k18ec138@dtu.ac.in

2022 2nd Asian Conference on Innovation in Technology (ASIANCON) | 978-1-6654-6851-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ASIANCON55314.2022.9909086

***Abstract*— Hepatitis C is a prominent viral infection affecting the liver, which has an estimated 58 million active cases throughout the world, with 1.5 million new infections adding to this tally every year. With a death toll of 290,000 by 2019, early diagnosis is the key to curing this infection with a recovery rate of 95%. In this paper, we aim to introduce an optimized XGBoost algorithm for predicting early progression to Hepatitis C which has been implemented on a labelled dataset with subjects of both genders and has been obtained from the UCI Machine Learning Repository. It contains data from 615 subjects aged between 19 and 77. The classification has been conducted primarily based on parameters such as Albumin (ALB), Alkaline Phosphate Enzyme (AKP), alanine aminotransferase (ALT), Bilirubin (BLT), Aspartate aminotransferase enzyme (AST), Cholesterol (CHOL), Serum cholinesterase (CHE) and gamma-glutamyl transferase (GGT).**

Our proposed methodology predicts progression towards Hepatitis C using XGBoost optimized using Monotonic Constraints and feature Interaction constraints. The prediction architecture is further optimized using regularization parameters and introducing boundary conditions such as Max-Depth, Learning rate and N-estimation. The resultant methodology predicts progression to Hepatitis C with an accuracy of 98.6%, which is higher when compared to other algorithms including Logistic Regression, LGBM, Decision Tree and SVM-RBF.

Keywords—Hepatitis C, Optimized XGBoost, Feature Interaction Constraint, Regularization, Learning Rate, N-estimation, Machine Learning

I. INTRODUCTION

Hepatitis C is a liver disease caused by the "Hepatitis C Virus"(HCV). This virus can cause chronic and acute Hepatitis, ranging from a mild illness that lasts a few weeks to a severe illness that lasts until death [1].HCV is a virus that can spread through the bloodstream. The most common types of infection are used on small amounts of blood [2]. It can spread due to drug injection, unsafe healthcare, transfusion of blood or blood plasma, unsafe injection practices and unsafe sexual practices. Early diagnosis becomes the key factor in treating HCV infection, whose recovery rate for early-stage treatment can go up to 95% using antiviral medications [3]. This recovery rate decreases exponentially as the disease progresses, thus this research aims at early diagnosis of progression to Hepatitis C.

Research and Advancements in Machine Learning have led to a significant increase in accuracy, precision, True-Positive rate etc. while conducting classification, regression and creating new predictive model architecture and this increase in accuracy is accompanied by a decrease in error rate. In this research, we intend to reduce dependency on a particular parameter while considering the diagnosis of disease and rather taking a holistic approach, wherein the machine itself determines the parameters to focus on and predict progression to Hepatitis C.

To create a predictive model for diagnosing the progression of Hepatitis C, XGBoost algorithm the has been used for classification [4][5]. Novelty in this paper includes optimizing the XGBoost algorithm using effective parameter selection, which has been implemented using Feature Interaction Constraint (FIC) and Monotonic Constraint. Optimizing further the prediction architecture involves optimizing Max Depth Learning Rate, N-Estimation and Col-Sample Bytree, along with manipulating regularization parameters such as alpha, lambda and gamma. The dataset used for this research contains data of 615 subjects of both gender and age between 19 and 77 suspected of suffering from Hepatitis C disease, which might further advance to liver cirrhosis and liver fibrosis. The results are then collected and the trained model is then implemented on a testing dataset, resulting in 98.4% accuracy, which is better than the accuracy obtained by implementing other algorithms such as Logistic Regression, LGBM, Decision Tree and SVM-RBF algorithm. The results are compared for further practical analysis and the Quantitative comparative efficiency of our model is obtained.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:17:31 UTC from IEEE Xplore. Restrictions apply.

The paper is organized as follows- Section 2 introduces the XGBoost algorithm; Section 3 Briefly introduces the process and methodology for XGBoost optimization using Regularization, boundary implementation and effective feature selection; Section 4 details the dataset and the parameters used for classification; Section 5 presents the algorithm proposed in this research paper; Section 6 presents results. Finally, Section 7 concludes with a discussion derived from this study and its applicability for future works.

II. XGBoost Classifier

XGBoost (eXtreme Gradient Boosting) is a fully open-source library, designed for advanced use of the Gradient Boosting framework. It focuses on the speed, flexibility, and performance of models. In terms of speed and classification performance, it is second to none. It is known to outperform many algorithms aimed at supervised learning.

It introduces improvement in the general Gradient Boosting It improves on the general Gradient Boosting Algorithm by minimising errors that usually occur in a gradient descent algorithm. Gradient descent, a repetitive development algorithm, calculates how far our estimate deviates from the actual result of a given data point. There are some prerequisites for XGBoost to perform optimally.

Specifically, scaling of numerical features is a must, and encoding of categorical features cannot be avoided. Weights are very important in XGBoost. Each independent feature is provided with a value called weight, which is then input to the tree, which further analyses the dataset to predict results. The value of weights, if wrongly predicted by the tree, is increased, and these variables are then fed into another decision tree iteratively. These individual classification/prediction architectures are finally combined to produce an accurate prediction model.

Objective Function-:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y_i}^{(t-1)} + f_t(\mathbf{x}_i)) + \Omega(f_t)$$

Real value (label) known from the training data-set

Can be seen as $f(x + \Delta x)$ where $x = \hat{y_i}^{(t-1)}$

XGBoost objective function analysis

Fig. 1. Objective function of XGBoost algorithm

Ref: medium.com – XGBoost https://dimleve.medium.com/xgboost-mathematics-explained-58262530904a

III. Optimization of XGBoost Algorithm

Optimizing the XGBoost algorithm increases the accuracy of classification as well as increases the accuracy of prediction of progression to Hep C. For Optimization of XGBoost, we have taken into account the following parameters-:

A. Boundaries Implementation

The boundaries that have been implemented for this research are as follows-:

1) Maximum tree Depth:

This parameter determines Each tree's maximum depth. A deep tree may improve performance, but it also increases complexity and opportunities for abundance. The value must be greater than zero. The default value is 6.

2) Learning Rate

As the model progresses toward its goal, the size of the step in each repetition is determined by the level of reading [8]. Low reading levels slow down calculations and necessitate more rounds to achieve the same reduction in residual error as a high-level learning model. However, it increases the chances of gaining better access. The value should be in the range of 0 to 1. The default value is 0.3.

3) N estimation

Determines the tree in our collection in total. It is equal to the number of development cycles. The value must be a number greater than 0. The default is 100.

4) Col-sample_Bytree

This parameter determines the sample or portion of a particular column that we take into consideration while classifying using XGBoost. This parameter is random, and the optimal value for classification has been procured by the repetitive iteration method. The value ranges from 0 to 1. Lower value might help tackle overfitting, but it also leads to less equity. The default value of this parameter is 1.

B. Regularization parameters

1) Alpha:

If you work with a large number of features, it may improve your performance [9]. It can be any whole number. The default is 0.

2) Lambda:

It may help to minimize overload. It can be any whole number. The default is 1.

3) Gamma:

Gamma, an artificial control parameter, depends on other parameters we are utilizing in the classification and prediction architecture. The higher the gamma, the higher the familiarity. It can be any whole number. The default is 0.

The above parameters have been implemented to optimize the XGBoost classification and prediction algorithm. The optimal value of the above parameters has been recorded and the optimal accuracy thus obtained has been calculated and presented in this research.

C. Feature Selection

Selecting the most significant hyper-parameters is a prominent determining factor for increasing the accuracy of the classification and our prediction models [10]. In this research, we have implemented the following approaches for effective parameter selection-:

1) Feature Interaction Constraint (FIC)

It allows users to specify which variables are allowed to interact with one another and which are not. Improves predictive performance as a result of focusing on effective interactions – whether through domain-specific knowledge or algorithms that rank interactions. Generates predictions with less noise; better generalization. Provides Greater user control over what the model can fit. For example, due to regulatory constraints, the user may wish to exclude some interactions even if they perform well.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:17:31 UTC from IEEE Xplore. Restrictions apply.

IV. *DATA-SET ANALYSIS* -:

Training data for a Machine learning algorithm is a major input for the algorithm to work at its optimum capacity. For our analysis, we have considered an HCV prediction dataset [11], as shown in Table 1. The dataset presents laboratory values of blood donors and Hepatitis C subjects' other features like age. The data was obtained from [11]. The dataset consists of 615 subjects aged between 19 and 77 years. The dataset contains people from both genders. The target attribute for classification is Category which gives us disease category. Disease categories are- hepatitis, fibrosis, and cirrhosis.

TABLE I. DATASET IMPLEMENTED FOR TRAINING AND DEVELOPING ARCHITECTURE FOR PREDICTING PROGRESSION TO HCV INFECTION [11]

	Category	Age	Sex	ALB	ALP	ALT	AST	BIL	CHE	CHOL	CREA	GGT	PROT
1	0=Blood Donor	32	m	38.5	52.5	7.7	22.1	7.5	6.93	3.23	106	12.1	69
2	0=Blood Donor	32	m	38.5	70.3	18	24.7	3.9	11.17	4.8	74	15.6	76.5
3	0=Blood Donor	32	m	46.9	74.7	36.2	52.6	6.1	8.84	5.2	86	33.2	79.3
4	0=Blood Donor	32	m	43.2	52	30.6	22.6	18.9	7.33	4.74	80	33.8	75.7
5	0=Blood Donor	32	m	39.2	74.1	32.6	24.8	9.6	9.15	4.32	76	29.9	68.7
6	0=Blood Donor	32	m	41.6	43.3	18.5	19.7	12.3	9.92	6.05	111	91	74
7	0=Blood Donor	32	m	46.3	41.3	17.5	17.8	8.5	7.01	4.79	70	16.9	74.5
8	0=Blood Donor	32	m	42.2	41.9	35.8	31.1	16.1	5.82	4.6	109	21.5	67.1
9	0=Blood Donor	32	m	50.9	65.5	23.2	21.2	6.9	8.69	4.1	83	13.7	71.3
10	0=Blood Donor	32	m	42.4	86.3	20.3	20	35.2	5.46	4.45	81	15.9	69.9
11	0=Blood Donor	32	m	44.3	52.3	21.7	22.4	17.2	4.15	3.57	78	24.1	75.4
12	0=Blood Donor	33	m	46.4	68.2	10.3	20	5.7	7.36	4.3	79	18.7	68.6
13	0=Blood Donor	33	m	36.3	78.6	23.6	22	7	8.56	5.38	78	19.4	68.7
14	0=Blood Donor	33	m	39	51.7	15.9	24	6.8	6.46	3.38	65	7	70.4
15	0=Blood Donor	33	m	38.7	39.8	22.5	23	4.1	4.63	4.97	63	15.2	71.9

TABLE II. PERCENTAGE DEPICTION OF DIFFERENT CATEGORIES

S No	Category	Percentage
1	Blood donors	86.67%
2	suspected blood donors	1.14%
3	Hepatitis patients	3.90%
4	Fibrosis patients	3.41%
5	Cirrhosis patients	3.41%

A. Parameter description

Dataset as shown in the figure contains the list of parameters taken into account while diagnosing a patient with Hepatitis C [12]. These parameters include age, sex, blood donor category, and various clinical terms described as follows-:

a) ALB

It determines how much albumin is in your blood. It ranges from 15 to 82.2.

b) ALP

It determines how much alkaline phosphatase enzyme is present in your bloodstream. It ranges from 11 to 416.

c) ALT

This gives us liver damage as hepatitis, cirrhosis, liver cancer, or other liver illnesses can all cause liver damage. It ranges from 1 to 325.

d) AST

This provides information regarding the Aspartate aminotransferase enzyme which is found in the liver and muscles. It ranges from 10 to 324.

e) BIL

BIL stands for the bilirubin test. The amount of bilirubin in the blood is measured by a bilirubin test. It ranges from 0.8 to 254.

f) CHE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:17:31 UTC from IEEE Xplore. Restrictions apply.

This provides information regarding the Serum cholinesterase enzyme whose level reflects the liver's synthetic function. It ranges from 1.4 to 16.4.

g) CHOL

It can tell you how much cholesterol and triglycerides are in your blood. It ranges from 1.5 to 9.67.

h) CREA

CREA stands for creatinine test. It is a test that determines how well your kidneys are filtering waste from your blood. It ranges from 8 to 107.9.

i) GGT

Many types of liver disease cause an increase in gamma-glutamyl transferase (GGT) activity. It ranges from 4.5 to 650.

j) PROT

It is a protein test. This measures the amount/presence of total protein in the blood sample in question. It ranges from 44 to 90.

For this study, another variable was introduced into the dataset. This variable, termed as "OUTPUT" variable, was selected as a binary variable, characterizing the subjects as 0/1, "0" for the subjects not suffering from hepatitis C and might be healthy, whereas "1" was assigned to subjects with a positive diagnosis of Hepatitis C infection. This new variable aided in classification methodology as well as was utilized in determining the error in *the* prediction architecture for HCV infection.

B. Exploratory Dataset Analysis-:

Analyzing the dataset is used to explore its main characteristics for determining significant and comparatively less significant characteristics.

1) Training dataset classification

As shown in Fig 2, out of 615 subjects, 533 were blood donors, 7 were suspected blood donors, 24 were suffering from hepatitis, 21 from fibrosis and 30 from cirrhosis. This provides us with the projection of the effectiveness of the dataset. Also, this analysis presents the severity of the situation that approximately 4% population is suffering from HCV Infection.

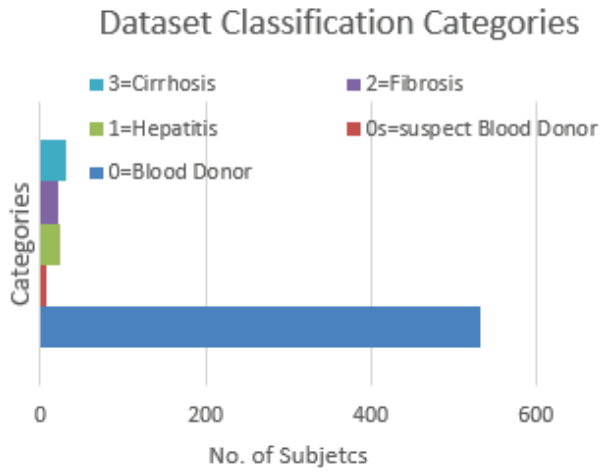


Fig. 2. Distribution of dataset among blood donors, suspected blood donors, hepatitis, fibrosis, and cirrhosis

2) Age v/s Diagnosis

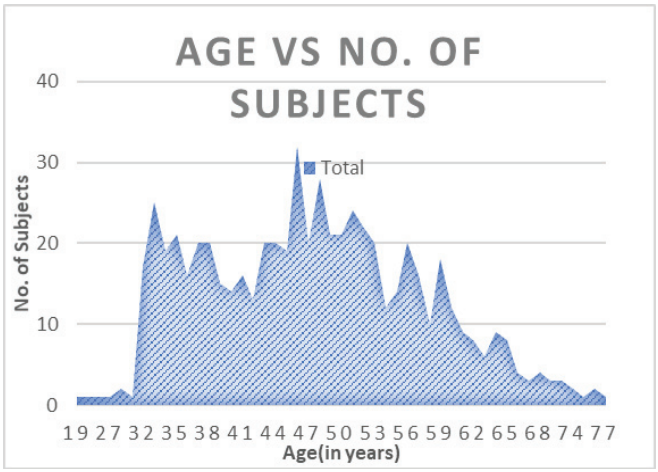


Fig. 3. Age-wise distribution of Subjects

The graph in Fig 3 depicts the disproportionality in the occurrence of HCV infection. The graph peaks at 47 years, but is still distributed fairly evenly between ages 32-59. It also shows that the dataset contains subjects of every age ranging from 19 to 77, thus increasing the reliability of the results obtained from the dataset.

3) Detailed Parameter Analysis for the dataset

A detailed analysis of all the parameters, including the total count of subjects, the mean of values, standard deviation, minimum value, 1st, 2nd and 3rd quartiles as well as their maximum value has been shown in Table 3. The minimum and maximum values are essential for providing the range of a particular parameter, which will be further utilized in scaling the parameters/features as a part of the data preprocessing step. Mean and standard deviation(std) are essential for determining the sanity of data, i.e. whether the data is spread or it is concentrated, whether there might be any outliers present or not etc. All these analyses have been carried out using Table 3.

4) Gender v/s Diagnosis

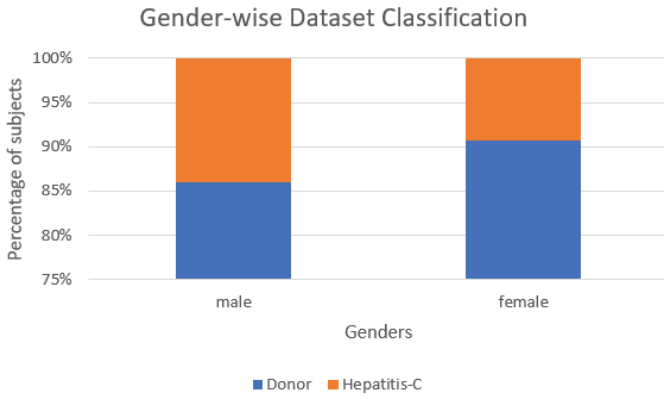


Fig. 4. Classification of infected subjects based on gender

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:17:31 UTC from IEEE Xplore. Restrictions apply.

TABLE III. PARAMETER DESCRIPTION AND ANALYSIS IN TERMS OF COUNT, MEAN, STD DEVIATION, MAXIMUM ETC.

	Unnamed: 0	Age	ALB	ALP	ALT	AST	BIL	CHE	CHOL	CREA	GGT	PROT
count	615.000000	615.000000	614.000000	597.000000	614.000000	615.000000	615.000000	615.000000	605.000000	615.000000	615.000000	614.000000
mean	308.000000	47.408130	41.620195	68.283920	28.450814	34.786341	11.396748	8.196634	5.368099	81.287805	39.533171	72.044137
std	177.679487	10.055105	5.780629	26.028315	25.469689	33.090690	19.673150	2.205657	1.132728	49.756166	54.661071	5.402636
min	1.000000	19.000000	14.900000	11.300000	0.900000	10.600000	0.800000	1.420000	1.430000	8.000000	4.500000	44.800000
25%	154.500000	39.000000	38.800000	52.500000	16.400000	21.600000	5.300000	6.935000	4.610000	67.000000	15.700000	69.300000
50%	308.000000	47.000000	41.950000	66.200000	23.000000	25.900000	7.300000	8.260000	5.300000	77.000000	23.300000	72.200000
75%	461.500000	54.000000	45.200000	80.100000	33.075000	32.900000	11.200000	9.590000	6.060000	88.000000	40.200000	75.400000
max	615.000000	77.000000	82.200000	416.600000	325.300000	324.000000	254.000000	16.410000	9.670000	1079.100000	650.900000	90.000000

As shown in Fig 4, Out of 615 subjects, 324 were identified as male subjects and 216 as Female subjects. Of the male subjects, 53 were suffering from Hepatitis-C, whereas among the female subjects, 22 were suffering from Hepatitis-C. This shows that males are more prone to Hepatitis-C disease. Our study can't properly justify whether males are more prominent to have HCV infection than women because of the sheer lack of data available. This study further requires extensive data collection and research.

V. METHODOLOGY IMPLEMENTED-:

The Proposed Optimized XGBoost model has been implemented using a series of interdependent steps (or phases) [13][14]. The initial phase consists of data manipulation and cleansing, in which the dataset was studied extensively and a unique parameter for each clinical parameter mentioned in the dataset was calculated to measure the significance of each parameter. The feature (parameter) of the dataset, with the least significance, was dropped by the algorithm from the dataset to reduce the complexity of the dataset and prevent over-fitting. The second phase included data cleaning, wherein missing values in the dataset were replaced using the mean and medians of other values present in the dataset by our proposed algorithm.

The third phase of the methodology involves conducting an extensive analysis of the clinical parameters included in our dataset, wherein the analysis of the classification of the dataset based on age, sex and number of scanning sessions were conducted and their effect on Hepatitis C progression was studied.

The fourth phase involves producing a correlation matrix for all the features present in the dataset as shown in Fig 5, to determine the correlation between different features to determine the least significant feature among others.

It is evident from the above correlation matrix (Fig 5) that cholesterol has a very high correlation with HCV Infection, indicating cholesterol is one of the prime indicators of HCV. The matrix also concludes a high correlation of AST with Age, concluding the fact that, Aspartate aminotransferase enzyme increases in our body as we age, hence AST was distinguished as a non-significant parameter.

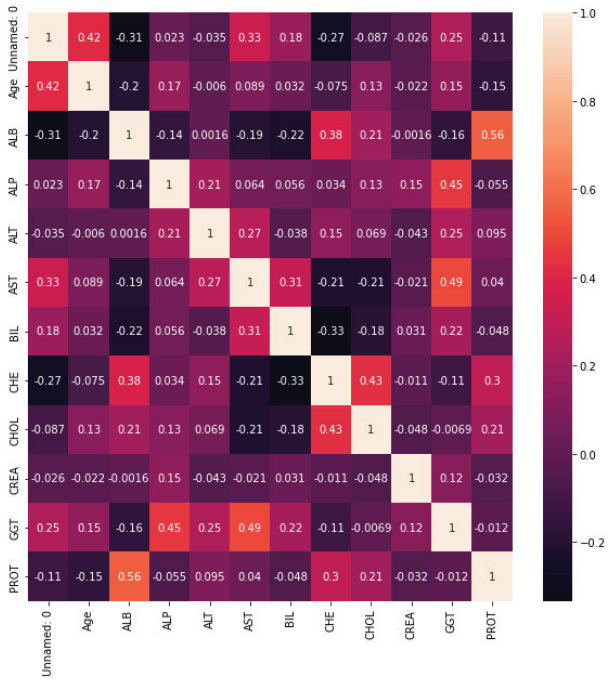


Fig. 5. Correlation Matrix

Further, the optimized XGBoost algorithm was implemented on the preprocessed data and the acquired results were compared with the corresponding accuracies of other algorithms implemented during this research. A flowchart of the complete methodology implemented as been shown in Fig 6.

A. *Algorithm Implemented*

1. Loading the Dataset and displaying the information.
2. Data interpretation with features.
3. Visualization of the data set.
4. Finding the target values and mapping correlation.
5. Preprocessing the dataset
6. Splitting the dataset into train and test sets.
7. Classify using XGBoost.
8. Optimizing the XGBoost using regularization parameters, boundary condition implementation and feature selection.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:17:31 UTC from IEEE Xplore. Restrictions apply.

9. Applying the randomized search cross-validation technique to find the optimum values of selected features.

10. Calculating the accuracy.

11. Performance evaluation.

12. Comparison with other classifiers.

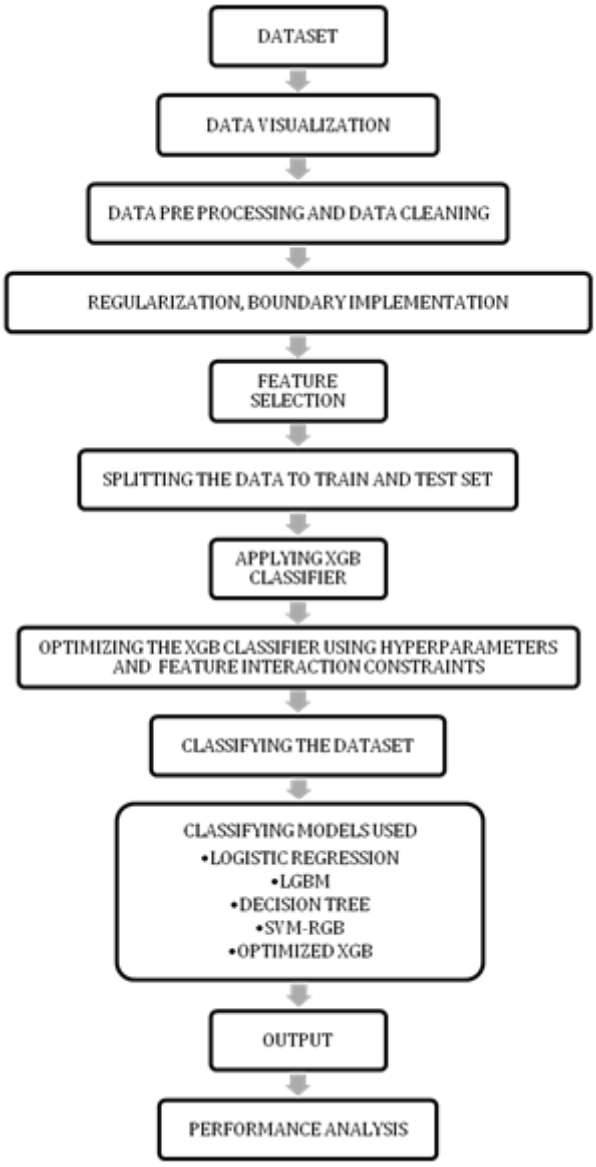


Fig. 6. Flowchart Depicting methodology Implemented

VI. RESULTS

The pre-processed dataset was further split into train and test sets for measuring the accuracy of our proposed model.

1) *Train-test split v/s Accuracy*

The model was evaluated for a train-test split of 0.8, 0.7 and 0.6. The proposed model was optimized at different test train split ratios and Random States (RS) to obtain the highest possible accuracy for classification as highlighted in Table 4.

TABLE IV. COMPARISON OF ACCURACY OF THE PROPOSED MODEL ON DIFFERENT TRAIN-TEST SPLIT RATIOS.

TRAIN- TEST SPLIT	***0.8***	***0.7***	***0.6***
(RS=1)	*0.95*	*0.95*	*0.96*
(RS=10)	*0.94*	*0.92*	*0.91*
(RS=42)	*0.95*	*0.95*	*0.95*
(RS=1)	*0.96*	*0.97*	*0.97*
(RS=10)	***0.99***	*0.95*	*0.96*
(RS=42)	*0.94*	*0.96*	*0.96*

2) *Performance Matrix*

The performance matrix of the Optimized XGBoost algorithm, is shown in Table 5, along with the performance matrix of other algorithms implemented on the same dataset. This provides us with a comparative analysis of our optimized XGBoost algorithm with other classification and prediction algorithms. The algorithms implemented for comparative analysis are-: Logistic Regression, Decision Tree, LGBM and SVM-RBF

TABLE V. PERFORMANCE MATRIX FOR THE OPTIMIZED XGB ALGORITHM

Model	**Precision**	**Recall**	**F1 Score**	**Accuracy**
Logistic Regression	0.89	0.87	0.83	0.83
Decision tree	0.85	0.83	0.89	0.89
LGBM	0.87	0.93	0.90	0.90
SVM-RBF	0.90	0.95	0.92	0.98
Optimized XGB	0.97	0.98	0.98	0.99

Each of the above-mentioned algorithms has been implemented on the same dataset [15]. A comparative study has been conducted extensively based on Precision, Recall, F1 Score and Accuracy (as shown in Table 6). The graph showcasing the accuracy of various algorithms implemented has been highlighted in Fig 7. The graph concludes that our proposed Neo-SVM model provides better accuracy than the algorithms under consideration when implemented on the same dataset.

The optimized XGBoost algorithm provides us with a signification prediction accuracy of 98.6%, as compared to Logistic Regression (83%), Decision Tree (89%), LGBM (90%) and SVM-RBF (97.8%).

Optimizing parameter FIS, selected ALB, AST, ALT, BIL, CHOL, PROT and Age to be the most significant parameters for classification. Other optimization techniques such as Gamma, lambda, Max-depth, learning rate and N-estimation were also implemented successfully for increasing the accuracy rate to 98.6%.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:17:31 UTC from IEEE Xplore. Restrictions apply.

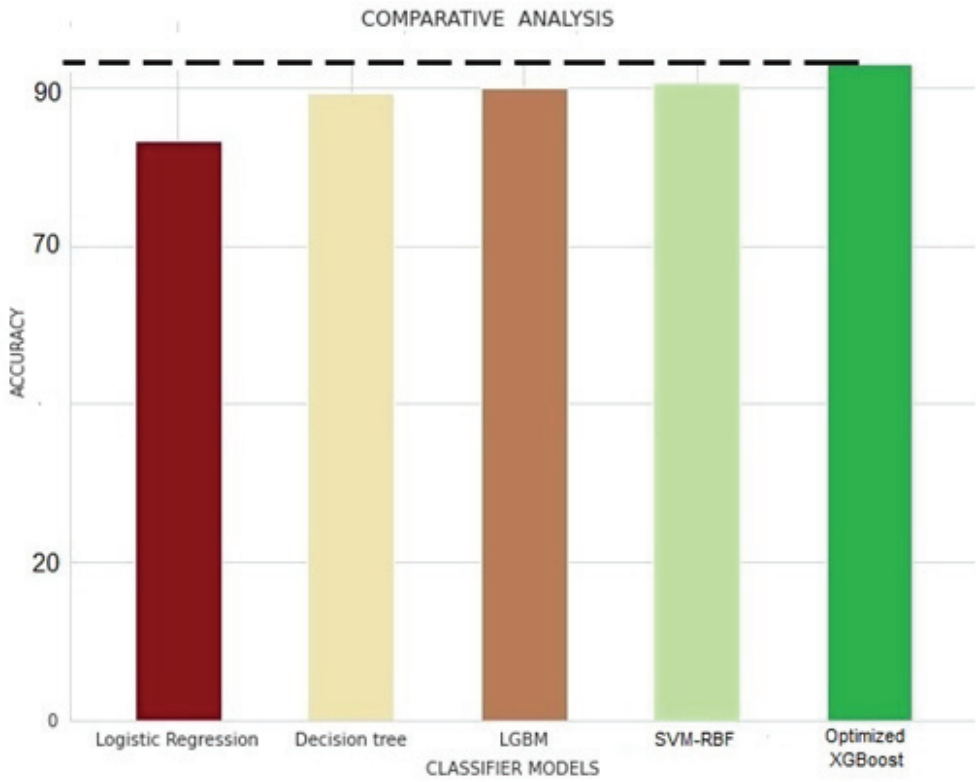


Fig. 7. Visualization of accuracies of various algorithms implemented on the dataset.

VII. Conclusion

The Optimized XGBoost model proposed in this research paper has been successfully implemented by efficient parameter selection and optimizing hyperparameters such as regularization parameters, boundary optimization and feature selection, which have been further evaluated based on recall, precision, f1-score and algorithm's performance as compared to the performance of other algorithms. From the study, It is concluded that HCV infection affects men and women disproportionately in different parts of the world. The proposed model for predicting progression to Hepatitis C infection provides a prediction accuracy of 98.6%, which is higher than the accuracy obtained from other algorithms when implemented on the same dataset.

VIII. Declaration of Conflicting Interests

There exists no conflict of interest. We, the authors of this research, hereby confirm that there are no known conflicts of interest associated with this publication. We also hereby confirm that there is no financial obligation or support provided for this publication.

IX. Author's Contribution

Kaustubh Ranjan Singh: Writing - Review & Editing, Supervision, Project administration;

Ritik Gupta: Methodology, Conceptualization, Validation, Formal analysis, Writing - Original Draft;

Ritik Kumar Kadian: Data Curation, Visualization, Software, Resources, Investigation;

Ritwik Singh: Formal analysis, Data Curation, Visualization;

X. Funding

No Funding was received for this work.

References

[1] Lauer, Georg M., and Bruce D. Walker. "Hepatitis C virus infection." New England journal of medicine 345.1 (2001): 41-52. https://www.nejm.org/doi/full/10.1056/nejm200107053450107

[2] Hepatitis C - Symptoms and causes https://www.mayoclinic.org/diseases-conditions/hepatitis-c/symptoms-causes/syc-20354278

[3] Hepatitis C Diagnosis and Tests for HCV https://www.webmd.com/hepatitis/diagnostic-tests-hepatitis-c-virus

[4] Chen, T., & Guestrin, C. (2016, August). xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM signed international conference on knowledge discovery and data mining (pp. 785-794). https://doi.org/10.1145/2939672.2939785

[5] Edeh Michael Onyema, Dalal Surjeet, Dhaou Imed Ben, Agubosim Charles Chuka, Smoke Chukwudum Collins, Richard-Nnabu Nneka Ernestina, Dahiya Neeraj "Artificial Intelligence-Based Ensemble Learning Model for Prediction of Hepatitis C Disease" on JOURNAL Frontiers in Public Health https://doi.org/10.3389/fpubh.2022.892371

[6] Gradient Boosting Classifier https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html

[7] Tianqi Chen, Tong He "xgboost: eXtreme Gradient Boosting" https://cran.microsoft.com/snapshot/2017-12-11/web/packages/xgboost/vignettes/xgboost.pdf

[8] Beginner's Guide to XGBoost for Classification Problems https://towardsdatascience.com/beginners-guide-to-xgboost-for-classification-problems-50f75aac5390

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:17:31 UTC from IEEE Xplore. Restrictions apply.

[9] Regularization and base learners in XGBoost https://campus.datacamp.com/courses/extreme-gradient-boosting-with-xgboost/regression-with-xgboost?ex=7

[10] Feature Importance and Feature Selection with XGBoost in Python https://machinelearningmastery.com/feature-importance-and-feature-selection-with-xgboost-in-python/

[11] HCV data Data Set https://archive.ics.uci.edu/ml/datasets/HCV+data

[12] Liver Function Tests https://www.ncbi.nlm.nih.gov/books/NBK482489/

[13] Hoofnagle, Jay H. "Course and outcome of hepatitis C." Hepatology 36.S1 (2002): S21-S29. https://pubmed.ncbi.nlm.nih.gov/12407573/

[14] L. Ma, Y. Yang, X. Ge, Y. Wan and X. Sang, "Prediction of disease progression of chronic hepatitis C based on XGBoost algorithm," in 2020 International Conference on Robots & Intelligent System (ICRIS), Sanya, China, 2020 pp. 598-601. https://dx.doi.org/10.1371%2Fjournal.pone.0208141

[15] Somaya Hashem, Gamal Esmat, Wafaa Elakel, Shahira Habashy "Comparison of Machine Learning Approaches for Prediction of Advanced Liver Fibrosis in Chronic Hepatitis C Patients" https://www.researchgate.net/publication/315844486_Comparison_of_Machine_Learning_Approaches_for_Prediction_of_Advanced_Liver_Fibrosis_in_Chronic_Hepatitis_C_Patients

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:17:31 UTC from IEEE Xplore. Restrictions apply.

PAPER • **OPEN ACCESS**

Analysis of Blended Concrete Cubes under Impact loading using ANSYS

To cite this article: Janhavi Singh and Shilpa Pal 2022 *IOP Conf. Ser.: Earth Environ. Sci.* **1084** 012067

View the article online for updates and enhancements.

You may also like

- Effects of Bioadditives and Commercial Additive on the Performance and Exhaust Emissions of a Gasoline Engine
Rico Aditia Prahmana
- Structure Formation and Characterization of PEMFC Catalyst Layers Blended with Multi-Walled Carbon Nanotubes
Takahiro Suzuki, Ryohei Hashizume and Masanori Hayase
- Development of the PDEODE-WEB Model in Blended Learning to Improve the Students Critical Thinking Skills
T. Mailani, Zulfarina and W. Syafii



This content was downloaded from IP address 14.139.251.98 on 25/10/2022 at 05:22

Analysis of Blended Concrete Cubes under Impact loading using ANSYS

Janhavi Singh[1] and **Shilpa Pal**[2]

Delhi Technological University, Delhi, India

Email: janhavisingh_2k20ste09@dtu.ac.in

Abstract. Blended concretes, with the partial replacement of the cement with secondary cementitious material like fly ash, rice husk ash, GGBS (Ground Granulated Blast Slag), silica fume etc. are gaining a wide range of applications in civil engineering nowadays. The behaviour of concrete structures has been studied by many researchers under different types of loads like impact load, wind load, earthquake load etc. Impact loading analysis has applications like ship impact resistance design for marine structures, impact-resistant structural design against military and terrorist attacks etc. This study aims to perform a FEM analysis of blended concrete cubes under impact load. Deformation and stress response has been obtained from the velocity impact simulations in ANSYS Explicit dynamics module. Parametric analysis has been done by changing the height of the impactor, shape of the impactor, boundary conditions and partial replacement of cement in the blended concrete with different supplementary cementitious material. It is observed that the strength of the blended concrete cube decreases with an increase in height of the impact.

1. Introduction

There is a growth in Portland cement demand in developing countries [1]. It is expected that the annual world cement production will shoot from 2540 million tonnes (2006) to around 3680-4380 million tonnes (2050) [2]. Industries' waste materials have the potential to be used in concrete as a replacement for aggregates or cement [1]. Some studies' results report that the cement can be replaced by various types of mineral admixtures [3]. Mineral admixtures such as metakaolin, GGBS (Ground Granulated Blast Slag) or Blast slag, Fly Ash, Rice Husk Ash, Palm Oil Fuel Ash and Silica Fume improve the quality of the concrete. They impart a significant influence on the hardened properties [4]. Partial replacement of the cement with such Supplementary Cementitious Materials (SCMs) brings down the consumption of cement and thus minimizes the impact on the environment [5]. Further use of such Supplementary Cementitious Materials helps to achieve a considerable amount of saving in cost and energy and promotes pollution reduction.[6]. So, it becomes clear that the blended concrete, in which the cement in the concrete is partially replaced by supplementary cementitious materials is gaining importance and the usage of such concretes should be extensively promoted in future. Thus there is a need to study such concretes under different types of loads. For our study, Fly ash and

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

Blast slag are used as supplementary cementitious materials, for partially replacing cement in the blended concrete. The cubes thus modelled in ANSYS are analysed for the impact loading. During the functional life of a structure or a structural element, there is a high probability of it being subjected to impact forces or repeated impacts. These forces are of various types and can occur due to a variety of reasons. For example- Moving vehicles may cause an impact accidentally on columns and walls [7]. Dropping of an object from height [8], water causing an impact on hydraulic structures, offshore structures or bridge piers [9, 10], can be some incidences when the impact force comes into play.

2. Specimen and configuration

Finite Element simulations of the 3 blended concrete cubes of size 150 mm x 150 mm x 150 mm have been performed using commercially accessible ANSYS 17.2 software. 2 Binary Blends and 1 Ternary Blend of concrete have been considered in which the cement is partially replaced with Fly ash and Blast slag. Table 1 shows the percentage with which the cement is replaced with the Fly ash and Blast slag in the 3 blends of concrete. The CRM (Cement Replacement Material)/Binder ratio as shown in Table 1 has been taken from the work of Jeong-Eun Kim et al. [11].

Table 1. CRM/Binder ratio in the 3 Blends of Concrete

Name of the Blend	Fly ash / Binder	Blast Slag / Binder
BB1(Binary Blend 1)	20%	0%
BB2(Binary Blend 2)	25%	0%
TB1(Tertiary Blend 1)	25%	25%

Table 2. Properties of Blended Concrete

Properties	BB1	BB2	TB1
Density (kg/cum)	2344	2349	2344
Young's Modulus (GPa)	26.3	25.5	26.8
Poisson's ratio	0.205	0.184	0.199
Bulk Modulus (GPa)	14.859	13.449	14.839
Shear Modulus (GPa)	10.913	10.076	11.176
Compressive Strength (MPa)	44.6	41.2	42.9

A spherical ball of diameter 130 mm and a steel box of base dimensions 100 mm X 100 mm have been used as impactors. Both the impactors have the same volume and are made with structural steel. Properties of the structural steel are given in Table 3. For this study Density (kg/cum), Young's Modulus(GPa), Poisson's ratio and Compressive Yield Strength (MPa) of the 3 blended concrete are obtained from the experimental work of Jeong-Eun Kim et al. [11]. These properties are specified in Table 3. The FE analysis has been performed under 3 different heights of impact (3m, 3.5m and 4m) for 3 different concrete cubes, with 2 boundary conditions - base fixed and all 5 sides of the concrete cube fixed.

Table 3. General material properties of Structural Steel

Properties	Values
Density (kg/cum)	7850
Young's Modulus(GPa)	200
Poisson's ratio	0.3
Bulk Modulus (GPa)	166.67
Shear Modulus (GPa)	76.92
Tensile Yield Strength (MPa)	250
Compressive Yield Strength (MPa)	250
Ultimate Tensile Strength (MPa)	460

3. Finite Element Modelling of Specimen

ANSYS 17.2 has been used for FE Analysis performed on the 3 blended concrete cubes. Generally, FE analysis consists of 3 steps [12]

Step 1 Modelling the Cube and then meshing
Step 2 Assignment of the boundary conditions required
Step 3 Performing the Analysis of the model and then extracting the solution (Analysis types, loading, and solving)

In this study, for each blended concrete, 2 models have been made (for spherical impactor and box impactor). For each of the 2 models, the analysis has been done for 2 different boundary conditions. In one of the conditions, the cube is made fixed at the base. In the other condition, the cube is fixed at the base and 4 sides. In total, 12 models have been made and they have been analyzed for 3 different heights of impact (3m, 3.5m and 4m). The impactor is considered to be falling freely under gravity from these 3 heights. When a ball falls freely from a height under gravity, the velocity with which it touches the ground is the square root of the 2 times the product of the acceleration due to gravity and the height of the free fall. For our study, we have taken the value of acceleration due to gravity as 9.8m/sq sec. 7.688 m/s, 8.282m/s and 8.854 m/s are the velocities of impactor obtained for the heights of impact as 3m, 3.5m and 4m respectively. The Analysis of these models has been performed and maximum equivalent von misses stress and maximum total deformation have been computed under a variety of impact load conditions.

3.1 Meshing of Specimen and Body Interaction

For this study, explicit meshing property and 20 mm mesh size have been used for concrete and steel impactor. An explicit dynamics model has been used for the impact load analysis in ANSYS. In explicit dynamics, the main parameter chosen were end time duration and velocity inputs. The meshing of the cube model is shown in Figure 1a and Figure 1b.

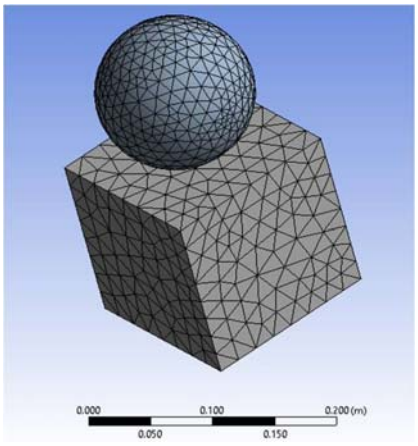


Figure 1a. FE Mesh- cube with sphere impact

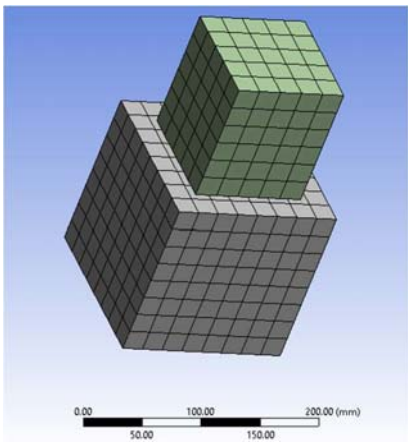


Figure 1b. FE Mesh-cube with box impact

3.2 Loading Conditions

Impact falling from varying heights was applied at the mid-point of the top surface of the cube. For the purpose of this study, the value of end time for loading is taken as 0.30 ms. The 3 velocity variations considered are 7.688 m/s, 8.282 m/s and 8.854 m/s.

4. Result and Discussion

4.1 Deflection and Stresses in the Blended Concrete under the Sphere shaped impactor

The maximum total deflection of cubes under impact was obtained aftersolving for the 3 blends.

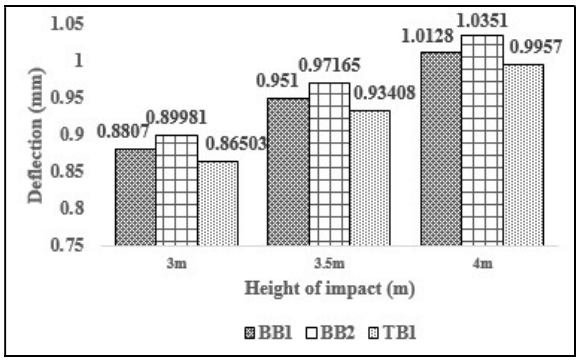


Figure 2a. Deflection- spherical impact- Base fixed

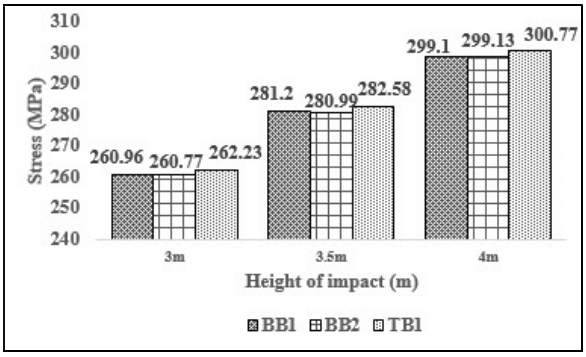


Figure 2b. Stress - spherical impact-Base fixed

Figure 2a and Figure 2b show the comparative results of deflection and stress, respectively, obtained on varying the height of impact for the 3 blended concrete cubes with the base fixed, when the shape of the impactor is a sphere. Deflection in the cubes fixed at the base, in BB2 is 2.07% to 2.2 % more than that in BB1 and that in TB1 is 1.688% to 1.77% higher than that in BB1. As it can be inferred from the graph, stresses developed in BB1 and BB2 are almost the same with a slight percentage difference of 0.07% to 0.01%. Stress developed in TB1 is 0.49% to 0.56% more than that in BB1. However, on fixing the 5 sides, the deflection in the cubes increases by 99.78% to 103.705% and the stress decreases by 5.57% to 9.928%. Figure 3 and Figure 4 show the comparison between the

deflection and stress respectively in the 3 Blended concretes, with the varying height of impact, under the sphere-shaped impactor when the base and the 5 sides are fixed respectively.

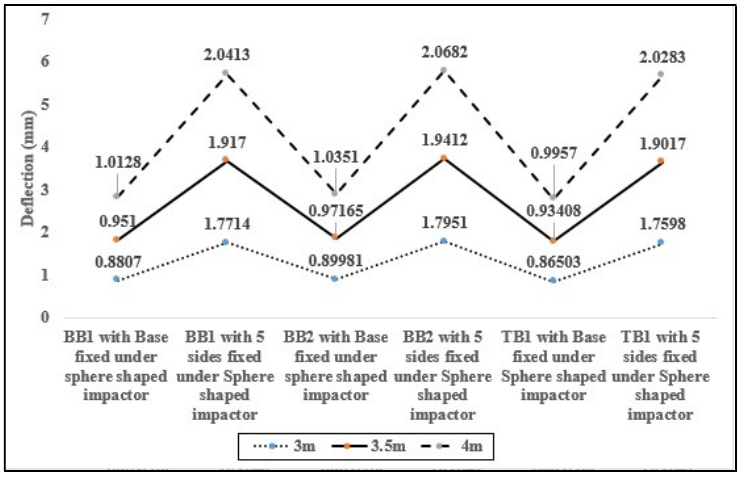


Figure 3. Comparison between the deflection under sphere-shaped impactor when the base and the 5 sides are fixed respectively

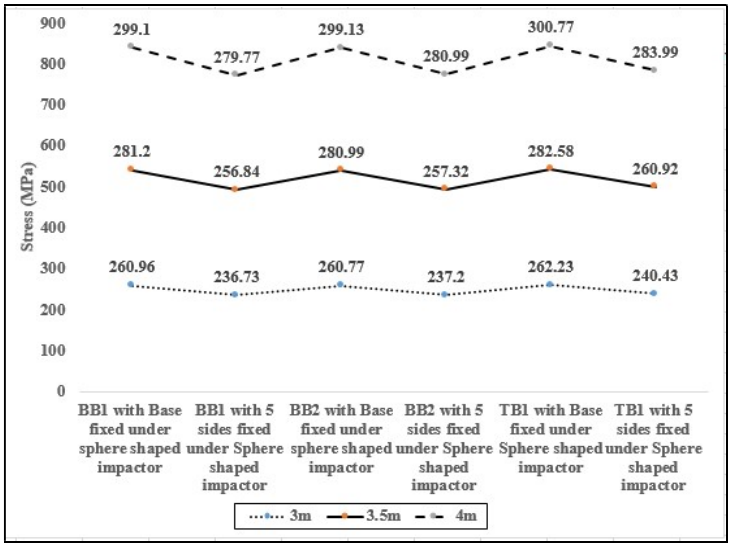


Figure 4. Comparison between the stress under sphere-shaped impactor when the base and the 5 sides are fixed respectively

4.2 Deflection and Stresses in the Blended Concrete under the Box shaped impactor

Under the Box-shaped impactor, the deflection in the cube, fixed at the base under different heights of impact in BB2 is about 2.5 % more than that in BB1 and that in TB1 is 1.67% lower than that in BB1. Stress developed in BB2 under different heights of impact is less than that in BB1, with a percentage difference of about 5.96%. Stress developed in TB1 under different heights of impact is about 3.31% more than that in BB1. However on fixing the 5 sides the deflection in the cubes

decreases by 66.61% to 66.62% for BB1, 69.62% to 69.63% for BB2 and 62.63% to 62.65% for TB1. The decrease in the stress is 77.05% to 77.07% for BB1, 82.60% to 81.63% for BB2 and 74.72% to 74.75% for TB1.

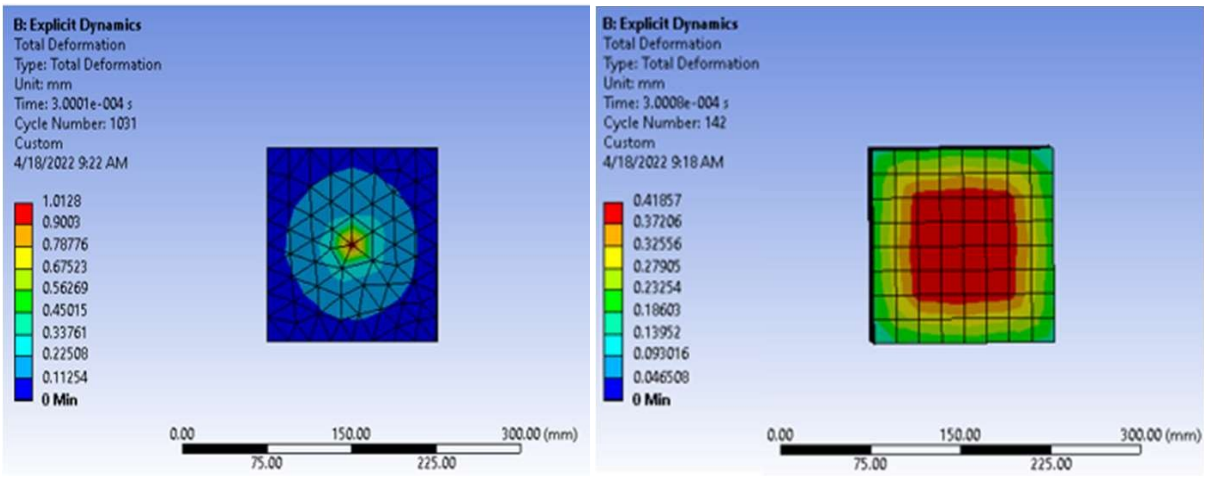


Figure 5. Deflection contours of BB1 cube under the height of impact 4m, under sphere-shaped impactor and box-shaped impactor

Deflection and stress contours of BB1 cube under the height of impact 4m, under sphere-shaped impactor and box-shaped impactor, are shown in Figure 5 and Figure 6 respectively. The maximum deformation and the maximum stress are obtained at the centre of the top surface of the cube.

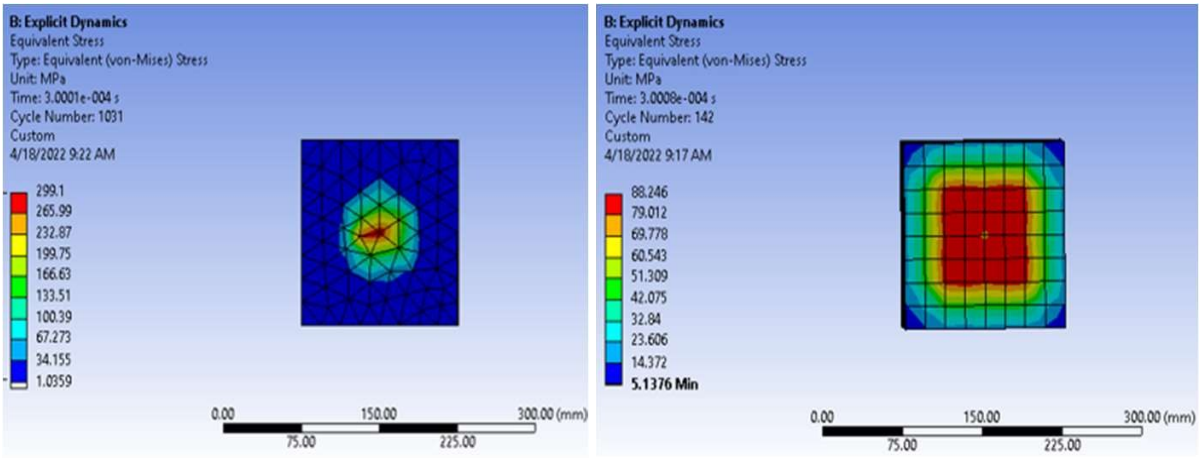
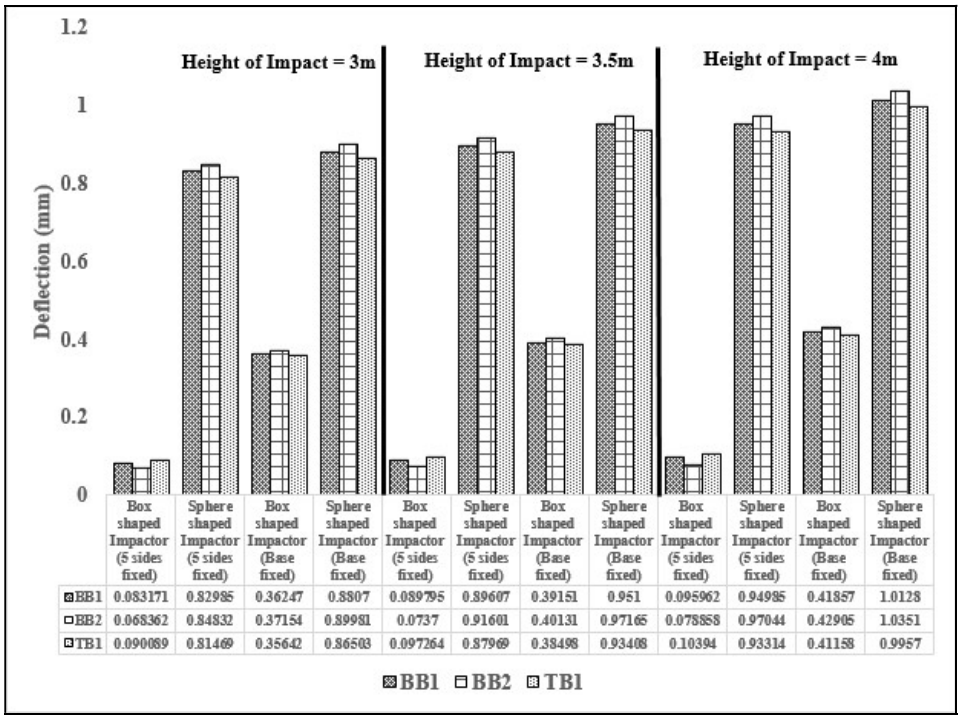


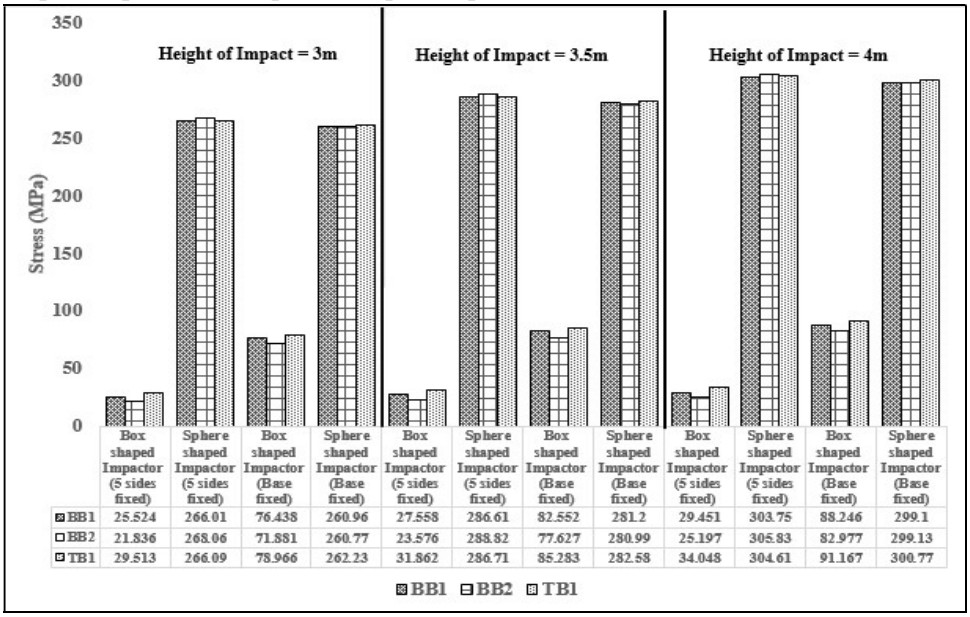
Figure 6. Stress contours of BB1 cube under the height of impact 4m, under sphere-shaped impactor and box-shaped impactor

Figure 7 and Figure 8 show the comparative analysis of the deflection and stress, respectively in the 3 blended concrete in various conditions, under Box-shaped impactor and Sphere-shaped impactor.



	Box shaped Impactor (5 sides fixed)	Sphere shaped Impactor (5 sides fixed)	Box shaped Impactor (Base fixed)	Sphere shaped Impactor (Base fixed)	Box shaped Impactor (5 sides fixed)	Sphere shaped Impactor (5 sides fixed)	Box shaped Impactor (Base fixed)	Sphere shaped Impactor (Base fixed)	Box shaped Impactor (5 sides fixed)	Sphere shaped Impactor (5 sides fixed)	Box shaped Impactor (Base fixed)	Sphere shaped Impactor (Base fixed)
BB1	0.083171	0.82985	0.36247	0.8807	0.089795	0.89607	0.39151	0.951	0.095962	0.94985	0.41857	1.0128
BB2	0.068362	0.84832	0.37154	0.89981	0.0737	0.91601	0.40131	0.97165	0.078858	0.97044	0.42905	1.0351
TB1	0.090089	0.81469	0.35642	0.86503	0.097264	0.87969	0.38498	0.93408	0.10394	0.93314	0.41158	0.9957

Figure 7. Comparative analysis of the deflection, in BB1, BB2 and TB1, for different heights of impact, in 2 different boundary conditions, under Box shaped impactor and Sphere shaped impactor



	Box shaped Impactor (5 sides fixed)	Sphere shaped Impactor (5 sides fixed)	Box shaped Impactor (Base fixed)	Sphere shaped Impactor (Base fixed)	Box shaped Impactor (5 sides fixed)	Sphere shaped Impactor (5 sides fixed)	Box shaped Impactor (Base fixed)	Sphere shaped Impactor (Base fixed)	Box shaped Impactor (5 sides fixed)	Sphere shaped Impactor (5 sides fixed)	Box shaped Impactor (Base fixed)	Sphere shaped Impactor (Base fixed)
BB1	25.524	266.01	76.438	260.96	27.558	286.61	82.552	281.2	29.451	303.75	88.246	299.1
BB2	21.836	268.06	71.881	260.77	23.576	288.82	77.627	280.99	25.197	305.83	82.977	299.13
TB1	29.513	266.09	78.966	262.23	31.862	286.71	85.283	282.58	34.048	304.61	91.167	300.77

Figure 8. Comparative analysis of the stress, in BB1, BB2 and TB1, for different heights of impact, in 2 different boundary conditions, under Box shaped impactor and Sphere shaped impactor

5. Result and Discussion

Using the finite element programme ANSYS, the impact behavior of the Blended concrete cubes with varying boundary conditions and shape of the impactor, under different impact velocities were analyzed. As a result of the findings, the following conclusions can be drawn:

- The deflection and the stress in the Blended concrete cubes increase with the increasing height of impact because as the height of impact increases, the impact energy increases.
- Cubes fixed on 5 sides show higher deflection and lower stresses. This is because less energy is dissipated when 5 sides of the cubes are fixed than when only the base of the cube is fixed.
- Less deflection and less stresses are developed when cubes are under impact of the Box shaped impactor than that of the spherical shaped impactor. This is because the area of application of the same impact increases in the former case.
- The Ternary Blended concrete comprising Blast slag and fly ash having developed higher stress and low deflection, shows better performance under impact load than the Binary blended concrete comprising only fly ash.

References

[1] Singh G and Bansal RS Effect of partial replacement of cement by silica fume and sand by quarry dust on strength and durability of concrete. 2017 Int. J. Eng. Res **4(9)** 414–418.

[2] Liu H, Bu Y, Nazari A, Sanjayan, JG and Shen, Z Low elastic modulus and expansive well cement system: the application of gypsum microsphere 2016 Constr Build Mater **106**, 27–34

[3] Bhargava GM and Kumar CJD State of art of usage of alternative materials in concrete 2015 Int. j. eng. **2(12)**,33–37

[4] Naveena SVD and Jagarapu DCK Experimental investigation on fibre reinforced beam column joint by partial replacement of cement with GGBS 2019 Int. j. recent technol **7(6)**, 215–219

[5] Dubey S, Chandak R and Yadav RK Experimental study of concrete with metakaolin as partial replacement of OPC 2015 Int. j. adv. eng. **2(6)**, 38–40

[6] Ahamed VS, Siddiraju S Study of strength of concrete with palm oil fuel ash as cement replacement 2016 Int. J. Civ. Eng. Technol **7(3)**, 337– 341

[7] Nili M and Afroughsabet V Combined effect of silica fume and steel fibers on the impact resistance and mechanical properties of concrete 2010 Int. J. Impact Eng. **37** 879–886

[8] Murali G and Ramprasad K A feasibility of enhancing the impact strength of novel layered two stage fibrous concrete slabs 2018 Eng. Struct. **175** 41–49

[9] Abid SR, Hilo A and Daek YH Experimental tests on the underwater abrasion of engineered cementitious composites 2018 Constr Build Mater **171** 779–792

[10] Abid SR, Shamkhi MS, Mahdi NS and Daek YH Hydro-abrasive resistance of engineered cementitious composites with PP and PVA fibers 2018 Constr Build Mater **187** 168–177

[11] Kim JE, Park WS, Jang Y, Kim SW, Sun-Woong Kim,Yi- Hyun Nam, Kim DG and Rokugo K Mechanical properties of energy efficient concretes made with binary, ternary, and quaternary cementitious blends of fly Ash, blast furnace slag, and silica fume 2016 Int. J. Concr. Struct. Mater. **10(3),** 97-108

[12] Armstrong CG Computer-Aided Design Butterworth-Heinemann Ltd 1994 (Department of Mechanical & Manufacturing Engineering, Queen's University of Belfast, Ashby Building, Stranmillis Road, Belfast BT9 5AH, UK) Volume 26

Analysis of Software Project Reports for Defect Prediction Using KNN

Rajni Jindal, Ruchika Malhotra and Abha Jain

***Abstract*— Defect severity assessment is highly essential for the software practitioners so that they can focus their attention and resources on the defects having a higher priority than the other defects. This would directly impact resource allocation and planning of subsequent defect fixing activities. In this paper, we intend to predict a model which will be used to assign a severity level to each of the defect found during testing. The model is based on text mining and machine learning technique. We have used KNN machine learning method to predict the model employed on an open source NASA dataset available in the PITS database. Area Under the Curve (AUC) obtained from Receiver Operating Characteristics (ROC) analysis is used as the performance measure to validate and analyze the results. The obtained results show that the performance of KNN technique is exceptionally well in predicting the defects corresponding to top 100 words for all the severity levels. Its performance is less for top 5 words, better for top 25 words and still better for top 50 words. Hence, with these results, it is reasonable to claim that the performance of KNN is dependent on the number of words selected as independent features. As the number of words increases, the performance of KNN also gets better. Apart from this, it has been noted that KNN method works best for medium severity defects as compared to the other severity defects.**

***Index Terms*—Receiver Operating Characteristics, Text mining, Machine Learning, Defect, Severity, K-Nearest Neighbour**

I. INTRODUCTION

NOW-a-days, various defect reporting/ tracking systems such as Bugzilla, CVS etc. are maintained for open source software repositories. These systems play an important role in tracking the defects which may be introduced in the source code [14]. These defects are then reported in a defect management system (DMS) for further analysis. Although, the issues of software are tracked using defect tracking systems which store the reported defects along with their details.

However, the data present in such systems is generally in unstructured form. Hence, text mining techniques in combination with machine learning techniques are required to analyze the data present in the defect tracking system.

Manuscript received March 14, 2014; March 31, 2014
Prof. Rajni Jindal is with Indira Gandhi Delhi Technical University for Women, Delhi, India (email: rajni.jindal@yahoo.co.in)
Dr. Ruchika Malhotra (Corresponding Author phone: 91-011-26431421) is with Delhi Technological University, Delhi, India (email: ruchikamalhotra2004@yahoo.com)
Abha Jain is with Delhi Technological University, Delhi, India (email: me_abha@yahoo.com)

In the present scenario, an automated tool is required to collect the data from software repositories so that it can be analyzed and interpreted in order to make generalized conclusions. The defects in the software may be associated with various severity levels. For instance, catastrophic defects are the most severe defects and a failure caused by such defects may lead to a whole system crash [1], [5].

In this paper, we mine the information from the NASA's database called PITS (Project and Issue Tracking System), by developing a tool that will first extract the relevant information from PITS using text mining techniques. After extraction, the tool will then predict the defect severities using machine learning techniques. The defects are classified into five categories of severity by NASA's engineers as very high, high, medium, low and very low. In this work, we have used K-nearest neighbor (KNN) technique to predict the defects at various levels of severity. The prediction of defect severity will help the researchers and software practitioners to allocate their testing resources on more severe areas of the software. The performance of the predicted model will be analyzed using Area Under the Curve (AUC) obtained from Receiver Operating Characteristics (ROC) analysis.

The rest of this paper is organized as follows: Section 2 reviews the key points of available literature in the domain. Section 3 describes the research method used for this study, which includes the data source and model evaluation criteria. Section 4 presents the result analysis. Section 5 concludes the paper and outlines directions for future work.

II. LITERATURE REVIEW

Nowadays, the analysis of defect project reports available in various open source software repositories has become the most essential step towards the successful completion of an error free software project. These defect reports are contained within the defect database and correspond to the defects which are encountered in the real-life systems. The defects occurring in such real-life systems are detected during testing by developer or anyone who is involved in the development of the product and are reported in a defect management system (DMS) or a bug tracking system. Later on, these defects are notified to the one responsible for the identification of its cause and its correction [18]. Each defect has its separate defect report which contains the detailed information about that defect. This information generally includes; ID of the defect, summary of the defect and associated severity of the defect. Till date, few authors have analyzed the defect project reports available in different open source software repositories

ISBN: 978-988-19252-7-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

for software defect prediction i.e. for predicting whether a A very effective tool based on Natural Language Processing (NLP) was developed by Runeson et al. [18] and Wang et al. [22] that was used to detect duplicate reports. Cubranic and Murphy [4] analyzed an incoming bug report and proposed an automated method that would assist in bug triage to predict the developer that would work on the bug based on the bug description. Canfora and Cerulo [3] discussed how software repositories can help developers in managing a new change request, either a bug or an enhancement feature. Also, a lot of empirical work has been carried out in predicting the fault proneness of classes in object-oriented (OO) software systems using a number of OO design metrics [2], [5], [7], [8], [11], [12], [15], [16], [23], [25]. Although, these studies were based on finding the relationship between OO metrics and fault proneness of classes, but did not focus on the severity of faults. Till date, there are only a few studies which were based on finding the relationship between OO metrics and fault proneness of classes at different levels of severity of faults.

The most efficient work in the field of fault severity has been done by the authors Singh et al. [21]. They have analyzed the performance of models at high, medium and low severity faults and found that the model predicted at high severity faults has lower accuracy than the models predicted at medium and low severities. The validation of the proposed models was done using various OO metrics like CBO, WMC, RFC, SLOC, LCOM, NOC, DIT on the public domain NASA dataset KC1 using DT and ANN as the machine learning methods and LR as the statistical method. The conclusion drawn was that DT and ANN models outperformed the LR model and that CBO, WMC, RFC and SLOC metrics are significant across all severity of faults and DIT metric is not significant across any severity of faults. LCOM and NOC are not found to be significant with respect to LSF. Somewhat same results were also concluded in the paper by Zhou and Leung [24]. They have investigated the fault-proneness prediction performance of OO design metrics with regard to ungraded, high, and low severity faults by employing statistical (LR) and machine learning (Naïve Bayes, Random Forest, and NNge) methods. From both the above papers, it was summarized that the design metrics are able to predict low severity faults in fault-prone classes better than high severity faults in fault-prone classes. Bayesian approach was also used by the author Pai [17] in his work to find the relationship between software product metrics and fault proneness. Shatnawi and Li [20] focused on identifying error-prone classes in post-release software evolution process. They studied the effectiveness of software metrics and examined three releases of the Eclipse project. They observed that, the accuracy of the prediction decreased from release to release and that there are only a few metrics which can predict class proneness in three error-severity categories.

The work proposed in this paper is similar to the work done by Menzies and Marcus [13]. The authors have presented an automated method named SEVERIS (SEVERity Issue assessment) which is used to assign severity levels to the defect reports by using the data from NASA's Project and Issue Tracking System (PITS). Their method is based on the automated extraction and analysis of textual descriptions from issue reports in PITS by using various text mining techniques. particular part of the software is defective or not.

They have used a rule learning method as their classification method to assign the features with proper severity levels, based on the classification of the existing reports. Similar work has also been done by Sari and Siahaan [19]. They have also developed a model for the assignment of the bug severity level. They have used the same pre-processing tasks (tokenization, stop words removal and stemming) and feature selection method (InfoGain), but, have used SVM as their classification method. Lamkanfi et al. [10] have also analyzed the textual description using text mining algorithms in order to propose a technique that is used to predict the severity of a reported bug against three open – source projects viz. Mozilla, Eclipse and GNOME using Bugzilla as their bug tracking system and Naïve Bayes as their classifier.

III. RESEARCH METHODOLOGY

In this section, we present our research method. We first introduce the data source which elaborates on the dataset being used in our study followed by the text classification framework that we have used in order to extract the relevant words from the defect descriptions. Finally, we describe our model evaluation criteria.

A. Data Source

We have collected the defect data from an open source NASA's dataset called PITS (Project and Issue Tracking System). There are various projects that come under PITS database all of which were supplied by NASA's Software Verification and Validation (IV & V) Program. We have used PITS B project wherein the data has been collected for more than 10 years and includes all the issues that have been found in the robotic satellite missions and human rated system. The focus of our study is to investigate the predictiveness of the model with regard to the severity of defects. Therefore, we are interested in the number of defects at each severity level as shown in Table I.

NASA's engineers have classified severity 1 defects as Very High, severity 2 defects as High, severity 3 defects as Medium, severity 4 defects as Low and severity 5 defects as Very Low. We are interested only in the last four severity levels i.e. severity 2, severity 3, severity 4 and severity 5 as it can be seen from table 1 that there are no severity one issues in the defect data. This is so because these defects are of very high severity level and therefore possibility of such defects in the software become very rare.

We went through this dataset and extracted the summary of each defect from all the reports. We then analyzed these textual descriptions and applied the text mining techniques to extract the relevant words from each report. At a later stage, machine learning method was used to assign the severity level to each defect based on the classifications of existing reports. As we know that the standard machine learning methods work well only for the data with fewer number of attributes. Therefore, before we can apply machine learning to the results of text mining, we have to reduce the number of words, referred to as the dimensions (i.e. attributes) in the data.

ISBN: 978-988-19252-7-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Hence, we applied different methods of text mining for dimensionality reduction in the following order: tokenization, stop word removal, stemming, feature selection and weighting.

TABLE I. NUMBER OF DEFECT REPORTS OF PITS B DATASET AT EACH SEVERITY LEVEL

	Severity 1	Severity 2	Severity 3	Severity 4	Severity 5
Pits B	0	23	523	382	59

1)Pre-processing

Pre-processing is the first and foremost step of text mining which is done in order to remove the irrelevant words from the document. Irrelevant words are the words which are not important for the learning task and rather their usage can substantially degrade performance of machine learning methods. The three most popular methods which are used for pre-processing are tokenization, stop words removal and stemming. Tokenization is the process of converting a stream of characters into a sequence of tokens. We have done tokenization by replacing the punctuation with blank spaces, removing all the non-printable escape characters and converting all the words to lowercases. Thereafter, all the stop words like prepositions, conjunctions, articles, common verbs, nouns, pronouns, adverbs and adjectives were removed from the dataset by using a list of English stop words. Finally, stemming was performed which removes words with the same stem and keeps the stem as the feature. For example, the words "train", "training", "trainer" and "trains" can be replaced with "train". All the words obtained after preprocessing were called as 'features'.

2)Feature selection

Even after performing a series of pre-processing tasks, the number of words in the document can still be very large. Therefore, feature selection method is used in order to further reduce the dimensionality of the feature set. There are a number of such methods available in the literature like document frequency, term frequency, mutual information, information gain, odds ratio, χ2statistic, term strength etc. These methods use an evaluation function that is applied to a single word. Thereafter, all these words are ranked by their independently determined scores, and then the top scoring words are selected.

In this paper, InfoGain measure is used to rank all the features obtained after pre-processing and then the top 'N' scoring features are selected based on the rank. According to the InfoGain measure, the best words are those that most simplifies the target concept, which is in our case, the distribution of severities [13]. Suppose a data set has 80% severity=5 issues and 20% severity=1 issues. Then that data set has a class distribution C_0 with classes c(1) = severity5 and c(2) = severity1 with frequencies $n(1) = 0.8$ and $n(2) = 0.2$. The number of bits required to encode an arbitrary class distribution C_0 is $B(C_0)$ defined as follows:

$$B(C) = -\sum_{c \in C} [{}^{n(c)}/_{N}] \log_2 [{}^{n(c)}/_{N}] \quad (1)$$

Where,

$$N = \sum_{c \in C} n(c) \quad (2)$$

If *A* is a set of attributes, then the number of bits required to encode a class after observing an attribute is:

$$B(C|A) = -\sum_{c \in A} \left[{}^{n(a)}/_{N}\right] \sum_{c \in C} p(c|a) \log_2 (p(c|a)) \quad (3)$$

Where,

$$p(c) = {}^{n(c)}/_{N} \quad (4)$$

The highest ranked attribute A_i is the one with the largest information gain; i.e., the one that most reduces the encoding required for the data after using that attribute; i.e.

$$Infogain(A_i) = B(C) - B(C|A_i) \quad (5)$$

3)Weighting and Normalizing

Now, these 'N' features can be represented as $t_1, t_2, \ldots, t_N$. The ith document is then represented as an ordered set of *N* values, called an N-dimensional vector which is written as $(X_{i1}, X_{i2}, \ldots, X_{iN})$ where X_{ij} is a weight measuring the importance of the jth term t_j in the *i*th document. The complete set of vectors for all documents under consideration is called a vector space model or VSM.

There are various methods which can be used for weighting the terms. We have used TFIDF approach for calculating the weights, which stands for Term Frequency Inverse Document Frequency. Term frequency is simply the frequency of the jth term, i.e. t_j, in document *i*. This value makes the terms that are frequent in the given document more important than the others. The second value i.e. inverse document frequency is given by $\log (n/n_j)$ where n_j is the number of documents containing term t_j and *n* is the total number of documents. This value makes the terms that are rare across the collection of documents more important than the others.

Now, before we use the set of N-dimensional vectors, we will first need to normalize the values of the weights. It has been observed that 'normalizing' the feature vectors before submitting them to the learning algorithm is the most necessary and important condition.

B. Model Evaluation

As we all know, defect-prone documents are treated as the positive instances in the context of defect prediction [9]. On this basis, we can categorize the defect prediction results into four different types as defined below:

- TP (True Positive): defect-prone documents that are classified correctly;

ISBN: 978-988-19252-7-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

- FN (False Negative): defect-prone documents that are wrongly classified to be defect- free;
- TN (True Negative): defect-free documents that are classified correctly;
- FP (False Positive): defect-free documents that are wrongly classified to be defect-prone.

To measure the performance of the predicted model, we have used the following performance evaluation measures:

1) Sensitivity

It measures the correctness of the predicted model and is defined as the percentage of the documents correctly predicted to be defect prone. It is also referred to as Recall. Mathematically we can define sensitivity as,

$$Sensitivity = \frac{TP}{TP + FN} * 100 \quad (6)$$

2) Receiver Operating Characteristics (ROC) analysis

ROC curve is defined as a plot of sensitivity on the y-coordinate versus its 1-specificity on the x- coordinates [6]. The main objective of constructing ROC curves is to obtain the required optimal cut-off point that maximizes both sensitivity and specificity. An overall indication of the accuracy of a ROC curve is the area under the curve (AUC). Values of AUC range from 0 to 1 and higher values indicate better prediction results.

3) Validation method used

The validation method used in our study is Hold-out validation (70-30 ratio) in which the entire dataset is divided into 70% training data and remaining 30% as test data. There are two methods by which we can assign the cases. Either we can randomly assign the cases based on relative number of cases or else we can use a partitioning variable. We have not randomly assigned the cases, but have rather used a partitioning variable to assign the cases. In other words, we have used a variable that splits the given dataset into training and testing samples in 70-30 ratio. This variable can have the value either 1 or 0. All the cases with the value of 1 for the variable are assigned to the training samples and all the other cases are assigned to the testing samples.

To get more generalized and accurate results, we have done validation using 10 separate partitioning variables. A single variable is used at a time on the basis of which the given dataset will be divided randomly in the ratio of 70-30. The corresponding training samples will be used by KNN in order to predict the model and the remaining testing samples will be used to validate the model. The same process is repeated for 10 runs corresponding to each partitioning variable.

IV. RESULT ANALYSIS

In this section, we have analyzed the results corresponding to the top-5, 25, 50 and 100 words in order to predict the best model that gives the highest accuracy using sensitivity, AUC and the cut-off point as the performance measures. Tables II-V present and summarize the results with respect to high, medium, low and very low severity levels.

TABLE II. RESULTS OF KNN FOR TOP-5 WORDS

	High Severity Defects			Medium Severity Defects			Low Severity Defects			Very Low Severity Defects		
Runs	AUC	Sens	Cut-Off	AUC	Sens	Cut-Off	AUC	Sens	Cut-Off	AUC	Sens	Cut-Off
1	0.495	00.0	0.214	0.544	66.0	0.357	0.550	59.0	0.214	0.493	10.5	0.214
2	0.496	00.0	1.000	0.548	66.7	0.357	0.556	43.3	0.357	0.495	16.7	0.214
3	0.568	14.3	0.214	0.555	37.5	0.500	0.581	58.9	0.214	0.528	17.6	0.214
4	0.559	12.5	0.214	0.520	48.0	0.500	0.554	53.7	0.214	0.473	04.8	0.214
5	0.495	00.0	0.214	0.538	40.1	0.500	0.564	58.1	0.214	0.492	10.5	0.214
6	0.498	00.0	0.214	0.564	50.7	0.500	0.550	49.6	0.214	0.437	05.0	0.357
7	0.568	14.3	0.214	0.475	32.4	0.500	0.512	56.0	0.214	0.464	05.9	0.214
8	0.582	16.7	0.214	0.596	47.3	0.500	0.584	62.0	0.214	0.489	09.1	0.357
9	0.561	12.5	0.214	0.515	43.1	0.500	0.531	55.8	0.214	0.477	04.2	0.214
10	0.628	33.3	0.214	0.545	58.9	0.357	0.539	52.4	0.214	0.542	20.0	0.214

TABLE III. RESULTS OF KNN FOR TOP-25 WORDS

	High Severity Defects			Medium Severity Defects			Low Severity Defects			Very Low Severity Defects		
Runs	AUC	Sens	Cut-Off	AUC	Sens	Cut-Off	AUC	Sens	Cut-Off	AUC	Sens	Cut-Off
1	0.596	25.0	0.214	0.708	70.1	0.357	0.646	39.3	0.357	0.533	21.1	0.214
2	0.691	40.0	0.214	0.739	73.7	0.357	0.697	48.6	0.357	0.595	33.3	0.214
3	0.620	28.6	0.214	0.682	69.4	0.357	0.650	46.8	0.357	0.548	23.5	0.214
4	0.795	62.5	0.214	0.726	71.1	0.357	0.669	51.2	0.357	0.587	28.6	0.214
5	0.659	33.3	0.214	0.729	70.4	0.357	0.681	51.3	0.357	0.566	26.3	0.214
6	0.684	40.0	0.214	0.696	70.9	0.357	0.653	51.3	0.357	0.632	40.0	0.214
7	0.701	42.9	0.214	0.690	69.9	0.357	0.666	45.7	0.357	0.575	29.4	0.214
8	0.590	20.0	0.214	0.708	63.2	0.357	0.639	49.6	0.357	0.530	25.0	0.214
9	0.699	44.4	0.214	0.754	71.4	0.357	0.692	51.6	0.357	0.567	25.0	0.214
10	0.822	66.7	0.214	0.738	67.4	0.357	0.677	50.9	0.357	0.573	25.0	0.214

TABLE IV. RESULTS OF KNN FOR TOP-50 WORDS

	High Severity Defects			Medium Severity Defects			Low Severity Defects			Very Low Severity Defects		
Runs	AUC	Sens	Cut-Off	AUC	Sens	Cut-Off	AUC	Sens	Cut-Off	AUC	Sens	Cut-Off

ISBN: 978-988-19252-7-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

1	0.698	50.0	0.214	0.746	70.7	0.357	0.709	49.6	0.357	0.701	52.6	0.214
2	0.678	40.0	0.214	0.762	73.0	0.357	0.762	57.9	0.357	0.663	44.4	0.214
3	0.682	42.9	0.214	0.758	65.0	0.357	0.668	47.6	0.357	0.674	52.9	0.214
4	0.582	25.0	0.214	0.798	75.0	0.357	0.739	47.2	0.357	0.834	76.2	0.214
5	0.813	66.7	0.214	0.810	75.7	0.357	0.757	54.7	0.357	0.681	47.4	0.214
6	0.632	30.0	0.214	0.754	70.9	0.357	0.697	53.0	0.357	0.712	55.0	0.214
7	0.754	57.1	0.214	0.796	71.0	0.357	0.739	58.6	0.357	0.675	47.1	0.214
8	0.564	20.0	0.214	0.748	70.1	0.357	0.708	53.8	0.357	0.660	50.0	0.214
9	0.742	55.6	0.214	0.728	67.9	0.357	0.681	50.8	0.357	0.562	25.0	0.214
10	0.648	33.3	0.214	0.759	70.2	0.357	0.670	57.3	0.357	0.721	55.0	0.214

TABLE V. RESULTS OF KNN FOR TOP-100 WORDS

	High Severity Defects			Medium Severity Defects			Low Severity Defects			Very Low Severity Defects		
Runs	AUC	Sens	Cut-Off	AUC	Sens	Cut-Off	AUC	Sens	Cut-Off	AUC	Sens	Cut-Off
1	0.734	50.0	0.214	0.751	72.1	0.357	0.698	54.7	0.366	0.713	57.9	0.214
2	0.687	40.0	0.214	0.779	67.8	0.357	0.749	60.7	0.357	0.665	50.0	0.214
3	0.692	42.9	0.214	0.775	72.5	0.357	0.710	86.3	0.214	0.705	58.8	0.214
4	0.610	25.0	0.214	0.785	75.0	0.357	0.699	50.4	0.357	0.813	76.2	0.214
5	0.822	66.7	0.214	0.795	78.3	0.357	0.724	49.6	0.357	0.680	47.4	0.214
6	0.689	40.0	0.214	0.749	70.3	0.357	0.675	77.4	0.214	0.707	50.5	0.214
7	0.775	57.1	0.214	0.781	75.0	0.357	0.721	55.2	0.357	0.638	41.2	0.214
8	0.732	50.0	0.214	0.778	74.7	0.357	0.722	51.2	0.357	0.547	27.3	0.214
9	0.732	50.0	0.214	0.760	72.2	0.357	0.708	54.9	0.357	0.600	41.7	0.214
10	0.732	50.0	0.214	0.776	67.4	0.357	0.742	65.7	0.357	0.716	53.3	0.214

From table II, it becomes clear that the performance of the model is consistent for all the severity levels when AUC is considered for evaluation. This is so because the values of AUC lie in the range of 0.4 to 0.6 for all the 10 consecutive runs, irrespective of the severity levels. However, there is a huge difference in the model prediction if we look at the sensitivity column. The maximum value of sensitivity for medium and low severity level is 66.7%, whereas its maximum value corresponding to high and very low severity level is just 33.3%. This trend suggests that model can predict medium severity defects better than the defects having either high or very low severity level when the number of words is less. As we increase the number of words to 25, we can see from table III that the prediction capability of the model is again the best for the medium severity defects with AUC in the range of 0.68 to 0.74 and is worst for very low severity defects with the maximum value of AUC being 0.6. Also, the sensitivity values for very low severity defects are just in the range of 21.1% to 40.0%. This trend is suggesting that there should be maximum focus on the defects having medium severity level when the number of words considered is nominal.

A similar kind of trend can be seen from table IV, where the model has performed exceptionally well in predicting the medium severity defects as values for both AUC and sensitivity are highest as compared to other severity levels. But, it is evident from table V, that performance of the model is consistent with respect to all the four severity levels when considering AUC as the performance evaluation measure. However, the sensitivity values indicate that the model predicted with respect to medium and low severity defects should be preferred over the model predicted with respect to high and very low severity defects.

But if we compare the performance of the model in terms of the number of words considered for model prediction, then there is a very clear indication of the fact that the model performs exceptionally well when top 100 words were considered, and that its performance drastically reduces when the number of words considered for classification is less. Performance of the model is worst when top 5 words were taken into account, is still better for top 25 and 50 words, and best for top 100 words. Hence, with these results, it is reasonable to claim that the performance of KNN is dependent on the number of words selected as independent features. As the number of words increases, the performance of KNN also improves. Apart from this, we have also observed that KNN method works best for medium severity defects as compared to the other severity defects.

V. CONCLUSION

Due to the widespread use of open source software repositories, the use of defect tracking systems has become inevitable. The issues of the software are tracked using such systems which store the defects along with their details. Such defects introduced in the software may be of varying severity levels ranging from mild to catastrophic. Thus, defect tracking systems play an important role to capture the defect data. However, a common issue in such systems is that they are useful for storing day-to-day information. Moreover, the information contained within such systems is generally of unstructured form.

Hence in this paper, text mining and machine learning techniques were used to analyze the defect data present in such systems. We have used text mining techniques to mine the information from the database and thereafter KNN machine learning method was used to predict the defect severities. The results were validated using NASA dataset available in the PITS database and were analyzed using Area Under the Curve (AUC) obtained from Receiver Operating Characteristics (ROC) analysis. The obtained results showed that the performance of KNN is exceptionally well in predicting the defects corresponding to top 100 words for all the severity levels. Its performance is poor for top 5 words, better for top 25 words and still better for top 50 words. Hence, with these results, we concluded that the performance of KNN is dependent on the number of words selected as independent features. Apart from this, we have also observed

ISBN: 978-988-19252-7-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

that KNN method works best for medium severity defects as compared to the other severity defects.

Although we analyzed only one project of PITS in our study so far, we believe that the research results can be generalized to the other projects available in the PITS database. So, our future work involves replication of this work in other projects of PITS.

REFERENCES

[1] K.K. Aggarwal, Y. Singh, A. Kaur and R. Malhotra, "Empirical analysis for investigating the effect of object-oriented metrics on fault proneness: A replicated case study," *Software Process: Improvement and Practice*, vol. 16, no.1, pp. 39-62, 2009.

[2] C. Catal and B. Diri, "A systematic review of software fault prediction studies," *Expert Systems with Applications*,vol.36, pp. 7346-7354, 2009.

[3] G. Canfora and L. Cerulo, "How Software Repositories can Help in Resolving a New Change Request," *Workshop on Empirical Studies in Reverse Engineering*, 2005.

[4] D. Cubranic and G.C. Murphy, "Automatic bug triage using text categorization," *Proceedings of the Sixteenth International Conference on Software Engineering and Knowledge Engineering*, 2004.

[5] K.E. Emam and W. Melo, "The Prediction of Faulty Classes Using Object-Oriented Design Metrics," *Technical report: NRC 43609*, 1999.

[6] K.E. Emam, S. Benlarbi, N. Goel and S. Rai, "A validation of object-oriented metrics," *NRC Technical report ERB-1063*, 1999.

[7] I. Gondra, "Applying machine learning to software fault-proneness prediction," *The Journal of Systems and Software*, vol.81, pp. 186-195, 2008.

[8] T. Gyimothy, R. Ferenc and I. Siket, "Empirical validation of object-oriented metrics on open source software for fault prediction," *IEEE Transactions on Software Engineering*, vol.31, no.10, pp. 897-910, 2005.

[9] Y. Jiang, B. Cukic and Y. Ma,"Techniques for evaluating fault prediction models," *Empirical. Software. Engineering*, vol.13, no. 15, pp. 561-595, 2008.

[10] A. Lamkanfi, D. Serge, E. Giger and B. Goethals, "Predicting the Severity of a Reported Bug," *7th IEEE working conference on Mining Software Repositories (MSR)*, pp. 1-10, 2010.

[11] R. Malhotra and Y. Singh, "On the Applicability of Machine Learning Techniques for Object- Oriented Software Fault Prediction," *Software Engineering: An International Journal*, vol.1, no.1, pp. 24-37, 2011.

[12] R. Malhotra and A. Jain, "Fault Prediction Using Statistical and Machine Learning Methods for Improving Software Quality," *Journal of Information Processing Systems*,vol. 8, no.2, pp. 241- 262, 2012.

[13] T. Menzies and A. Marcus, "Automated Severity Assessment of Software Defect Reports," *IEEE International Conference on Software Maintenance (ICSM)*, 2008.

[14] G. Myers, T. Badgett, T. Thomas and C. Sandler, "The Art of Software Testing," second ed., John Wiley & Sons, Inc., Hoboken, NJ, 2004.

[15] N. Ohlsson, M. Zhao, M and M. Helander , "Application of multivariate analysis for software fault prediction," *Software Quality Journal*, vol.7, pp.51-66, 1998.

[16] H. Olague, L. Etzkorn, S. Gholston and S. Quattlebaum, "Empirical validation of three software metrics suites to predict fault-proneness of object-oriented classes developed using highly iterative or agile software development processes," *IEEE Transactions on Software Engineering*, vol.33, no.8, pp. 402-419, 2007.

[17] G. Pai, "Empirical analysis of software fault content and fault proneness using Bayesian methods," *IEEE Transactions on Software Engineering*, vol. 33, no. 10, pp. 675-686, 2007.

[18] P. Runeson, M. Alexandersson and O. Nyholm, "Detection of Duplicate Defect Reports Using Natural Language Processing," *29th IEEE International Conference on Software Engineering (ICSE)*,pp. 499 – 508, 2007.

[19] G.I.P. Sari and D.O. Siahaan, "An attribute Selection For Severity level Determination According To The Support Vector Machine Classification Result," *Proceedings of The 1st International Conference on Information Systems For Business Competitiveness (ICISBC)*, 2011.

[20] R. Shatnawi and W. Li, "The effectiveness of software metrics in identifying error-prone classes in post-release software evolution process," *The Journal of Systems and Software*,vol. 81, pp. 1868-1882, 2008.

[21] Y. Singh, A. Kaur and R. Malhotra,"Empirical validation of object-oriented metrics for predicting fault proneness models," *Software Quality Journal*, vol.18, pp. 3-35, 2010.

[22] X. Wang, L. Zhang, T. Xie, J. Anvik and J. Sun, "An Approach to Detecting Duplicate Bug Reports using Natural Language and Execution Information," *Association for Computing Machinery*, 2008.

[23] P. Yu, T. Systa, and H. Muller, "Predicting fault-proneness using OO metrics: An industrial case study," *In Proceedings of Sixth European Conference on Software Maintenance and Reengineering*, Budapest, Hungary, pp.99-107, 2002.

[24] Y. Zhou, and H. Leung,"Empirical Analysis of Object-Oriented Design Metrics for Predicting High and Low Severity Faults," *IEEE Transactions on Software Engineering*, vol. 32, no. 10, pp. 771-789, 2006.

[25] Y. Zhou, B. Xu, and H. Leung,"On the ability of complexity metrics to predict fault-prone classes in object -oriented systems," *The journal of Systems and Software*, vol.83, pp.660-674, 2010.

ISBN: 978-988-19252-7-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:10/October-2022 Impact Factor- 6.752 www.irjmets.com

ANALYZING AND PRIORITIZING RISKS INCLUDING CIRCULAR ECONOMY INITIATIVES IN SUPPLY CHAINS

Shweta Kumari*1, Akanksha Singh*2

*1,2Economics, Delhi Technological University, India.

DOI : https://www.doi.org/10.56726/IRJMETS30453

ABSTRACT

The continual use of resources is a key ingredient in the sustainability of thecompany circular economy helps to reduce the proper pressure on the environment, improving the security of the supply of raw materials, increasing competitiveness, stimulating innovation, boosting economic growth but also they impose risk on the industry. This paper works on the insights of supply chain (SC) system in India using statistical methods to understand the cause-effect of different risk associated with the supply chain with the help of further analysis using regression and averaging the combined effect of the risk on every other risk to prioritize the factors to reduce the risks.

I. INTRODUCTION

Circular economy highlights a closed system where the remaining amount of resource residue is heavily used. Previously, the circular economy was tested based on systemic and theoretical methods, but now many industries are looking at the possible implementation of circular economy in their processes. A well-developed sales force determines the organization's stability in its business community. Similarly, a well-integrated chain forms the basis for building a closed system to reduce wastage and leakage of energy and to eliminate solid waste and to address unnecessary resource use.

Risk analysis becomes a more focused and widely accepted activity among organizations hoping to ensure a smooth transition. To identify these risks and their implications, this paper refers to a number of studies and methods of evaluating / analyzing and proposing a risk prioritization strategy. For effective management of potential risks, Supply Chain management / staff need tobe equipped to take effective measures to reduce those risks. To do so, Supply Chain management / staff need to identify, prioritize and analyze all potential risks. This study aims to contribute to the Circular Supply Chain literature by addressing the prioritization and impact of risks on each other individually and in an integrated manner.

The study is organized into 5 phases. Section 2 provides a curtailed review of the literature on circular economy. Section 3 sets out the distinct approaches used in this study. Section 4 deals with the results and gives compact discussions while the conclusions are presented in sections 5.

II. LITERATURE REVIEW

In recent years, the Circular Supply Chain and Circular Economy have attracted the attention of businesses, policy makers, and academics, in improving the efficiency of Supply Chain. Many Organizations have established sustainable programs to reduce the various risks to the Supply Chain. Circular Economy assists organizations to acquire Circular Supply Chain's for waste management purposes and energy. Reason after Circular Economy can apply to various departments such as procurement, production and information systems within an organization. It requires organizations to rebuild the Supply Chain network around sustainable development. It also helps to identify hazards and minimize them with good ongoing practices. In [1] Recent research highlights the importance of risk analysis but has no definitions of how an organization can use such skills to determine corporate sustainability. The literature is reviewed on two fronts, namely, risk analysis in Circular Economy programs in Supply Chain and risk assessment in Supply Chain in the context of Circular Economy programs. The paper gives the cause and effect of each factor on the other factors associated with the supply chain.

[2] Manages incomplete, vague and uncertain information effectively by using neutrosophic values. [3] Highlights that polices of European countries favoring the circular economy should focus on the [12] subgroup of SMEs. [4] Examined 2823 European Union that highlights the role of capital intensity as the relationship between labour productivity and [11] environmental performance depends on it. [5] The study analyses the

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:10/October-2022 Impact Factor- 6.752 www.irjmets.com

relationship among interests of stakeholders, reputational risk and extent of corporate responsibility. [6] The authors reviewed 125 papers and proposed a systematic assessment of supply chain resilience. [7] The study analyzed barriers to the circular economy of Europe that are lack of resources, expertise, administration and legal procedure. [8] The authors discloses that all circular economy areas of action (take, make, distribute, use, and recover) of small and medium-sized enterprises are correlated to economic performance, but only make and use are related to environmental and social performance. [9] The authors focuses on the importance of procurement of resources from two sources for effective cost management. [10] The relations between dynamic capabilities and corporate sustainability are important as it provides an over view that sensing, seizing, and reconfiguring dynamic capabilities are good for CE strategies.

Table 1. List of Risks

Risk Categories	Risks	Acronym
Operational risks	Material delay	O1
	Worker' coordination	O2
	Material quality	O3
	Safety measures	O4
	Specified machine	O5
Financial risks	Supplier performance	F1
	Product quality	F2
	Marketing strategies	F3
	Controlled Cash-flow	F4
	Return on investment	F5
Experimental risks	Natural disaster	E1
	Processing environment	E2
	Machine maintenance	E3
Strategic risks	Branding	S1
	Reputed customers	S2
	Balanced offers	S3
	Vision statements	S4
	Report governing	S5
Quality risks	Customer satisfaction	Q1
	Product performance	Q2
	Product service life	Q3
	Resale value	Q4
Reputational risks	Genuine publicity	R1
	Decision making	R2
	Effectivity	R3
	Communication	R4
	Social responsibilities	R5
Compliance risks	Standards	C1

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:10/October-2022 Impact Factor- 6.752 www.irjmets.com

	Transparent process	C2
	Policies and procedures	C3
	Cargo thefts	C4

III. METHODOLOGY

A combination of calculation and analysis methods is adopted using numerical solution algorithms and analytical models. The risk matrix in Table from [1] is used for the inspection of the factors that are stiffly influencing the other factors. The mean and standard deviation of the population of the risk matrix on which cause and effect is studied and through which cause and effect arises are judged to view the mean and deviation of the population risk factors.

$$x = \Sigma R.F/n \quad for\ each\ row$$

Where R.F is Risk Factor in the row

$$y = \Sigma C.R.F/n\ for\ each\ column$$

Where C.R.F is Causation on Risk Factor for each column

2

$$\text{And } std = \sqrt{\Sigma(xi - x)\ /n}$$

Where n are the total no. of Risk factors the x and y mean was later analyzed by giving them ranks based on their magnitude of influence on the factors and getting influenced factors according to their ranks and the same was done for the standard deviations on both the cause and effect factors. The table is checked for the correlation for the total relation matrix whether there is any Correlation which shows the dependence of one risk over another. The standard deviation and the mean of the R.F and the C.R.F is been experimented for linear relationship by trying to see the different hyperplanes which fits the model to predict the mean of a sample given its dispersion to the mean which can help to analyze the risk dependence on other risks more accurately. Spearman Rank Correlation is used between the two given variables to measure the strength and direction of association between two ranked variables. It helps us by giving a monotonic function to describe the two variables.

IV. RESULTS

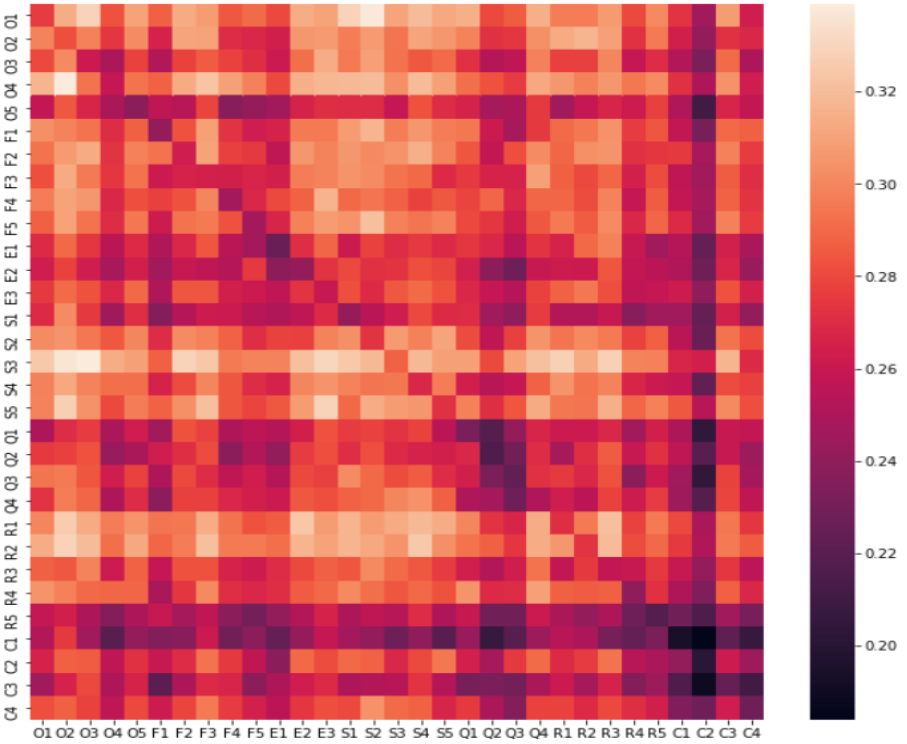


Figure 1. Heat map of Total Relation Matrix

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:10/October-2022 Impact Factor- 6.752 www.irjmets.com

A heat map is drawn to give a visual representation of how one risk is dependent on the other. A darker square denotes a weak relation while a lighter square denotes a strong relation. It is often helpful to draw a heat map to try to get an overview on which risks toconcentrate.

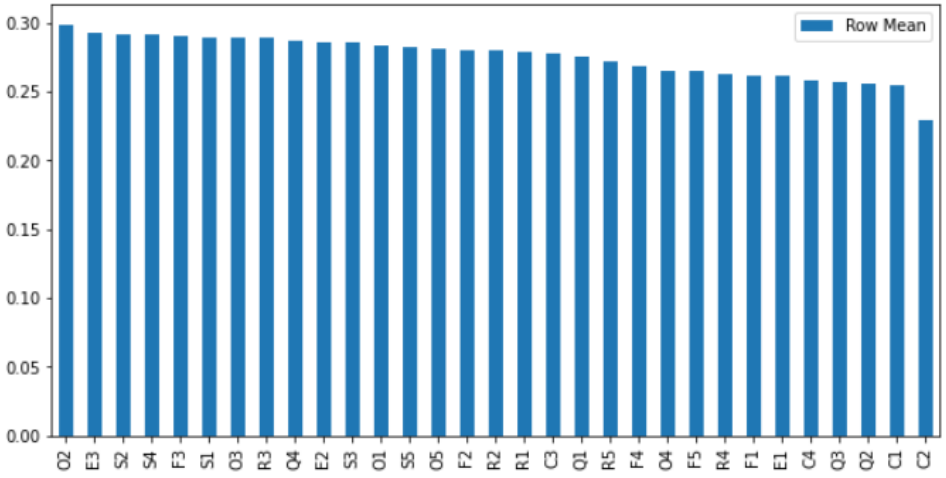


Figure 2. Mean of Risk Features on which cause and effect is studied

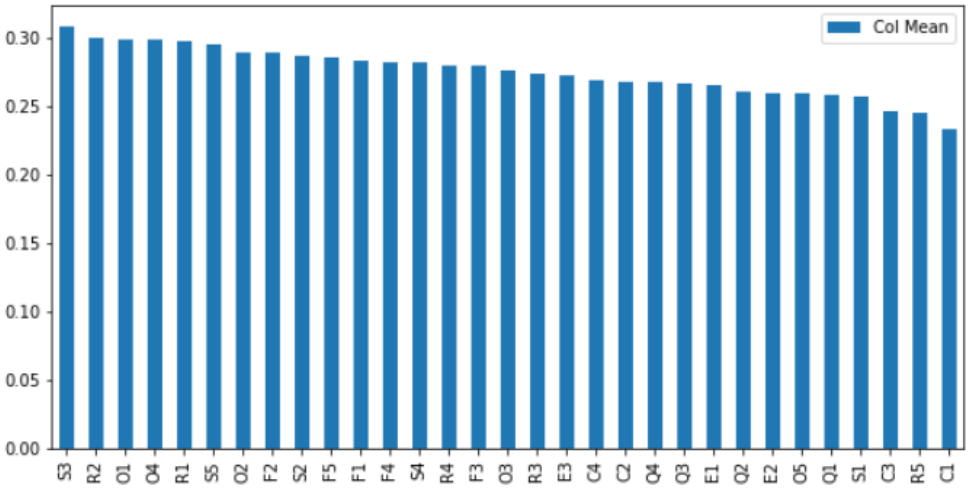


Figure 3. Mean of Risk Features through which cause and effect is studied

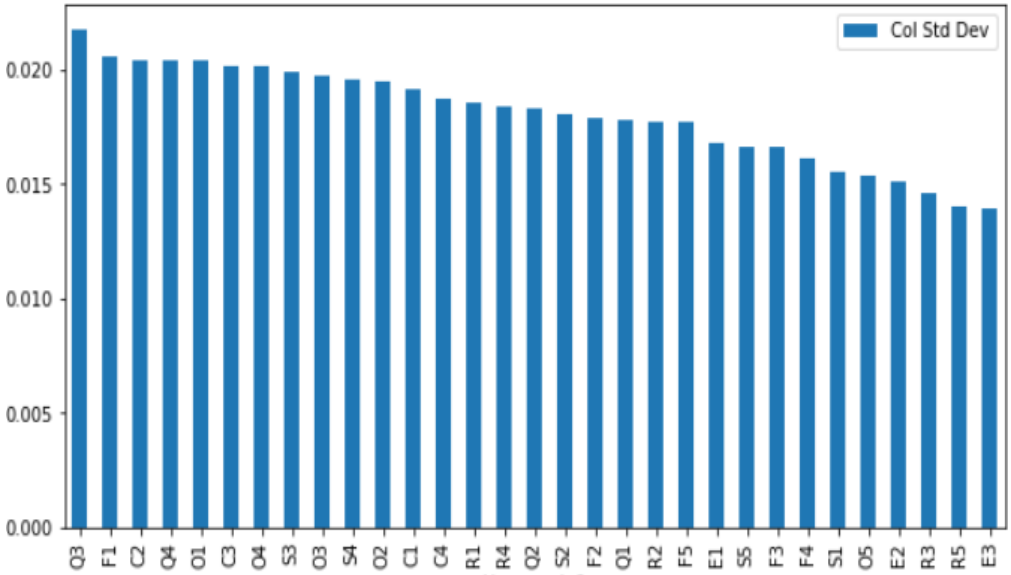


Figure 4. Standard Deviation of Risk Features through which cause and effect is studied

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:10/October-2022 Impact Factor- 6.752 www.irjmets.com

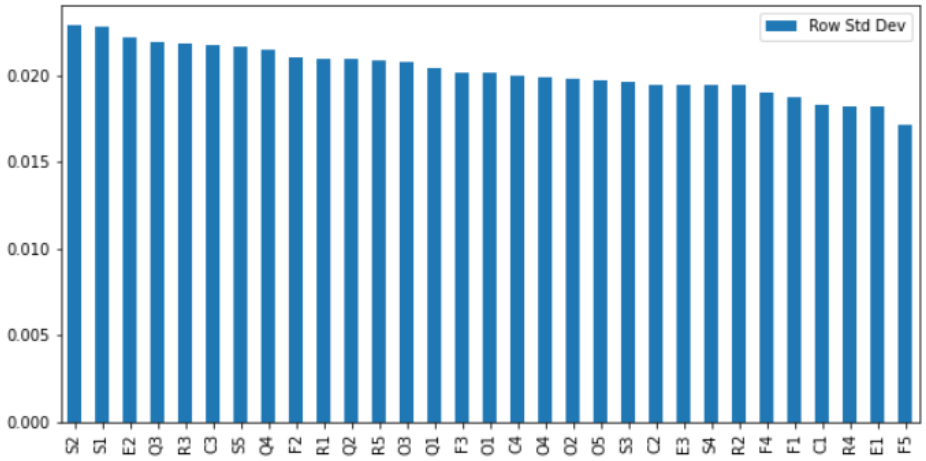


Figure 5. Standard Deviation of Risk Features on which cause and effect is studied

Figure 2 and 3 discuss the impact of one risk on all the other identified risks and ranks these risks based on their total average value. This can be used to understand the cause and effect of risks on one another to prioritize the importance of the risks. To further enhance the understanding of the risks, their standard deviations have been studied in Figure 4 and 5 which helps understand the volatility associated with them.

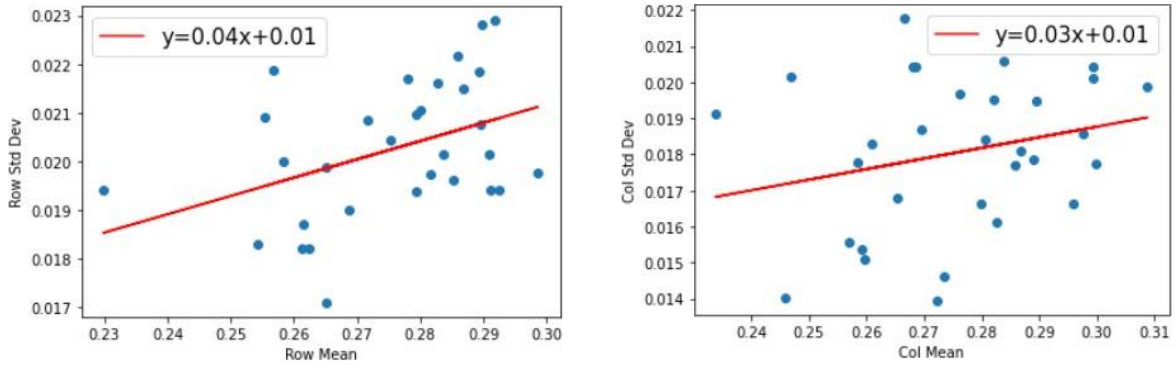


Figure 6. Linear Relationship of the standard deviation and mean on the risks

By comparing and plotting the mean and standard deviation of the given risks, and using regression analysis, it is concluded that there is no direct correlation between the two. It is not feasible to generalize the impact of one risk on all the others together. This can be understood as different organizations have different business environment risks associated with them.

Table 2. Spearman Correlation on Combination of Risks Statistics

	Correlation	Pvalue
Row Mean, Column Mean	0.2544354838709676	0.16719377687570175
Row Standard Deviation, Column Standard Deviation	0.07056451612903226	0.7060160474637789
Row Mean, Row Standard Deviation	0.3544354838709678	0.050423141750356254
Column Mean, Column Standard Deviation	0.22419354838709676	0.22533320460270215

Spearman Rank Correlation methodology was used to find the relation between the average values and their volatility for the risks when grouped together. A higher value indicates a stronger relation. Here all values are less than 0.4 which indicate a very weak relation. We cannot use this method to compare the risks when their impact is grouped. P-value fails to give accurate data for a smaller dataset and hence is ignored here.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:10/October-2022 Impact Factor- 6.752 www.irjmets.com

Table 3. Effective Scores of the 7 major risks identified as whole

Risk	Cumulative Rank inRow i.e. Cause	Cumulative Rank in Column i.e. Effect	Effective Score
Operational	1,7,12,14,22	3,4,7,16,26	11.2
Experimental	2,10,26	18,23,25	17.33
Financial	5,15,21,23,25	8,10,11,12,15	14.5
Strategic	3,4,6,11,13	1,6,9,13,28	9.4
Reputational	8,16,17,20,24	2,5,14,17,30	15.3
Compliance	18,27,30,31	19,20,29,31	25.65
Quality	9,19,28,29	21,22,24,27	22.25

Since studying individual risks have not revealed any pattern, they have been grouped together under their respective headings according to rank. Then the average of these ranks are taken to find an effective score to rank the 7 risks according to increasing order of value. A smaller value indicates a risk which should be of more importance. Here from Table 4. We can rank the risks inthe following order of priority.

Strategic > Operational > Financial > Reputational > Experimental > Quality > Compliance

V. CONCLUSION

This study reveals that seven major risks have been influenced within the Supply Chain; namely, operational, finance, environment, quality, reputation, compliance and strategic risks. This has a significant impact on the planning, processing, distribution and retrieval of the three companies that participated in the study. A small risk in one organization will affect the process of another organization because it is built under a single Circular Economy context. It is also known that every Supply Chain risk has its own impact on each other. Therefore, this paper examines therelationship between risks and influences effectively; with an analytical tool that defines the scope of influence systematically both individually and cumulatively.

This research work can be further enhanced by using numerous large datasets of different environments where the risk associated with each other can be studied. A simple case studycannot be used to generalize the risks of different circular economies. One risk itself cannotinfluence all the other risks, so managers need to go into the intricacies and understand the impact of one on the other. A wholesome overview of the major 7 risks can help managers commence the measures to safeguard the interest of organization in terms of sustainable development. This paper can be further implemented on the datasets of countries which have a similar environment to that of India, like Brazil, Indonesia or South Africa.

VI. REFERENCES

[1] Ethirajan, M, Arasu M, T, Kandasamy, J, K.E.K, V, Nadeem, SP, Kumar, A. Analysing the risks of adopting circular economy initiatives in manufacturing supply chains. Bus Strat Env. 2021; 30: 204– 236. https://doi.org/10.1002/bse.2617

[2] Abdel-Basset, M., Gunasekaran, M., Mohamed, M., & Chilamkurti, N. (2019). A framework for risk assessment, management and evaluation: Economic tool for quantifying risks in supply chain. Future Generation Computer Systems, 90(1), 489-502.

[3] Bassi, F., & Dias, J. G. (2020). Sustainable development of small and medium-sized enterprises in the European Union: A taxonomy of circular economy practices. Business Strategy and the Environment, 1–14. https://doi.org/10.1002/bse.2518

[4] Lannelongue, G., Gonzalez-Benito, J., & Quiroz, I. (2017). Environmental management and labour productivity: The moderating role of capital intensity. Journal of environmental management, 190, 158-169.

[5] Cassano, R. (2019). Corporate global responsibility and reputation risk management. Symphonya. Emerging Issues in Management, (1), 129-142.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:10/October-2022 Impact Factor- 6.752 www.irjmets.com

[6] Centobelli, P., Cerchione, R., & Ertz, M. (2020). Managing supply chain resilience to pursue business and environmental strategies. Business Strategy and the Environment, 29(3), 1215–1246.

[7] García-Quevedo, J., Jové-Llopis, E., & Martínez-Ros, E. (2020). Barriers to the circular economy in European small and medium-sized firms. Business Strategy and the Environment, 29(6), 2450-2464.

[8] Dey, P. K., Malesios, C., De, D., Budhwar, P., Chowdhury, S., & Cheffi, W. (2020). Circular economy to enhance sustainability of small and medium-sized enterprises. Business Strategy and the Environment, 29(6), 2145-2169.

[9] Inderfurth, K., & Kelle, P. (2011). Capacity reservation under spot market price uncertainty. International Journal of Production Economics, 133(1), 272-279.

[10] Khan, O., Daddi, T., & Iraldo, F. (2020). Microfoundations of dynamic capabilities: Insights from circular economy business cases. Business Strategy and the Environment, 29(3), 1479-1493.

[11] Perey, R., Benn, S., Agarwal, R., & Edwards, M. (2018). The place of waste: Changing business value for the circular economy. Business Strategy and the Environment, 27(5), 631-642.

[12] Prieto-Sandoval, V., Ormazabal, M., Jaca, C., & Viles, E. (2018). Key elements in assessing circular economy implementation in small and medium-sized enterprises. Business Strategy and the Environment, 27(8), 1525-1534.

Rendiconti Lincei. Scienze Fisiche e Naturali
https://doi.org/10.1007/s12210-022-01107-3

RESEARCH PAPER

Appraisal of hydrochemistry and suitability assessment for water in an agriculture-dominated aquatic ecosystem of Rajasthan, India

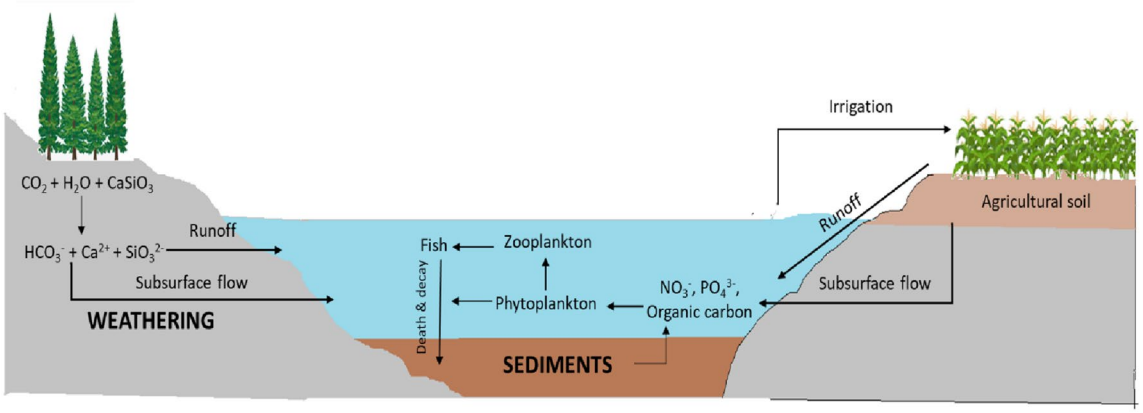
Chitrakshi[1] · **Anil Kumar Haritash**[1]

Received: 2 June 2022 / Accepted: 17 September 2022
© The Author(s), under exclusive licence to Accademia Nazionale dei Lincei 2022

Abstract

The study is the first comprehensive investigation of the water quality and geochemistry of an ecologically important large endorheic lake located in Rajasthan, India. The water quality and its suitability appraisal classified the water as fit for drinking, irrigation, and industrial application with the possible problem of corrosion. Most of the dissolved ions were observed to be coming from crustal/natural origin, with the exception of chloride. The water was observed to be $MgHCO_3$ type and CaMgCl type based on the relative abundance of ions. Considering the environmental relevance towards eutrophication, the Nitrogen:Phosphorous (N:P) value of 10:1 in water classified the lake in transition from oligotrophic to eutrophic with an urgent need of intervention to arrest addition of domestic wastewater and agricultural runoff. Additionally, Fourier Transmission Infra-Red (FTIR) investigation of sediments and soil revealed that the chemical signatures of soil and sediments are similar, and most of the sediments are transported into the lake from a short distance. The geochemical classification of water and mineralogical profile of soil and sediments of the lake confirmed that rock–water interaction (silicate weathering in particular) is controlling the water chemistry via direct cation exchange. Considering the importance of lake as a hotspot for tourists, migratory birds, crocodile conservation, aquaculture and its application for irrigation and drinking, present study recommends effective management of water quality for its sustainable use.

Graphical abstract



Keywords Eutrophication · Geochemistry · Langelier Saturation Index (LSI) · Sodium adsorption ratio (SAR) · Siliserh lake

✉ Anil Kumar Haritash
akharitash@dce.ac.in

Extended author information available on the last page of the article

Published online: 22 October 2022

1 Introduction

Rivers and lakes are important sources of fresh water. Whereas most of the human activities are centered around rivers resulting in poor water quality (Haritash et al. 2008; Kaushik et al. 2009), the lakes are mostly fresh water bodies and provide purer water than the river systems (Shan et al. 2020). There have been various studies on deducing the quality of groundwater (Kumar et al. 2017; Chitrakshi and Haritash 2018; Haritash et al. 2018; Ersoy and Karaca 2019; Aggarwal et al. 2021), but investigations on lakes are fewer. The primary source of lake water is precipitation which is the purest form of water that may enter the lake as runoff from surrounding areas or through small drains. Sometimes, lakes and ponds get formed naturally through some springs that enter natural depression and get collected. Healthy lakes not only supply the water for domestic, agricultural, and industrial uses but also provide number of environmental benefits influencing the quality. A healthy lake ecosystem maintains healthy natural and social balance at its location. Lakes replenish the groundwater, preserve the biodiversity and habitat of that area, support pisciculture, provide means of easy transportation and communication, provide moderating effect on climate, provide prime opportunities for recreation and tourism, and some lakes are used for extraction of salts and minerals (Haritash et al. 2017). The diverse activities supported by lakes thus help strengthen the economy as well. In the present scenario, the lakes in urban areas are under constant threat of degradation water quality because of rapid urbanization, industrialization, and indiscriminate use of agrochemicals (Bayram and Onsoy 2015; Ambastha and Haritash 2021). There are several sources of pollutants that could affect the quality of water resources (Karaca et al. 2017). Population growth, unplanned expansion of city area, urbanization, more requirement of land for agriculture, and lack of awareness among local people are the factors which affect a lake ecosystem adversely. Considering the important functions in different hydrological processes and environmental services rendered by lakes, man-made urban lakes and other natural wetlands have earned the environmental relevance over recent years. Many of the lakes have disappeared due to illegal filling, dumping of waste, and drying up of their catchment areas which have been reclaimed for uses like urbanization (Yadav et al. 2015). Apart from it, most of the lakes in India are under the menace of nutrient enrichment; which is responsible for algal blooms and subsequent choking up of eutrophicated waterbody (Haritash et al. 2015; De Donno et al. 2017; Pham et al. 2022). Since the fresh water sources are under threat and gradually becoming polluted, regular monitoring has become an important concern worldwide (Karaca et al. 2017; Rossi et al. 2019). Monitoring of physico-chemical and biological parameters of lake water may provide early warning signs which helps in detecting the contaminant inputs and sites. This also helps in determining the changes in quality over the time and helps to shape public policy before their after-effects become prominent (Goyal et al. 2021).

Keeping in view the facts mentioned above, the present study was undertaken at Siliserh lake which is an important historical (since 1845) man-made waterbody of Rajasthan state. Rajasthan is the largest state in India with an area of 342,239 km^2 (around 10% of total geographical area of India), and about 50% of the total area of the state (170,000 km^2) is covered by the great Indian desert. The population density of the state is more than 200 persons per km^2 (2011 census). The state is prone to droughts and events of scarcity of water. In a period of 102 years (1868–1969), the state has experienced nine (09) events of severe drought throughout the state, and 90 events of water scarcity in one or the other district of the state. The state has 58% arid, 36% semi-arid, and 6% of the total land area as sub-humid (Banerji and Upadhyay 1976). The subsurface geology of Alwar is a mix of rocks belonging to pre-Cambrian and Delhi subgroup. The deposits are chiefly large quartzite with schist, marble, and slate. Alwar district of Rajasthan lies in semi-arid zone with annual average precipitation of 595 mm, and some of the dry region receives rainfall as low as 100 mm per year. The district has faced 18 events of water scarcity (1868–1969), and the recent (2003–2014) decadal variation shows receding water table at about 85% of the locations. The average depth to water table is more than 40 m with annual fluctuation of more than 4.0 m (CGWB 2016). Considering the limitation of water resources, most of the cities in Rajasthan have natural or man-made lakes in or around the cities. Since the major source of water is precipitation which is limited to monsoon season (July to September), the quantity of water is significantly reduced during winters and pre monsoon. In recent years, there have been studies on deteriorating water quality in such lakes, but most of the studies are confined to bigger lakes in major cities (Kazmi et al. 2013; Rawal et al. 2014). The relatively less popular lakes like Siliserh lake in Alwar city are less studied in respect of its water quality (Vashistha et al. 2016). Further, there have been no studies for identifying the sources of pollution or geochemical attributes which affect the water quality in smaller lakes. Siliserh lake is one such less popular lake located in foothills of the Aravalli ranges. This region receives annual average rainfall of 675 mm and the

groundwater table is 40 m below ground level (mbgl). Therefore, the population of this region is dominantly dependent on Siliserh lake for water.

Since the lake receives agricultural runoff, domestic sewage, and the waste from bovine cattle from adjoining villages, it becomes imperative to monitor the water quality and identify the possible sources of contamination. Considering the utilization of water in different sectors, the present study was undertaken to assess the water quality and suitability of Siliserh lake for drinking, irrigation, and industrial use; to identify the sources which adversely affect the water quality and to understand the contribution of natural sources, viz. weathering, over the quality of water.

2 Materials and methods

2.1 Study area

Siliserh lake is located at 27º02′38″ North latitude, 74º2′79″ East longitude and is surrounded by Doba, Ringaspura, Kishanpura, and Paitpur villages. The lake is in the foothills of the Aravalli mountain range, at an altitude of 268 m above the sea level. The lake was formed by constructing a dam which is 1025 m long to prevent/collect runoff from the catchment. The catchment area of this lake is 11.2 km^2 and the culturable command area is 7.2 km^2. The full tank level and top bank level of the lake are 9.29 m and 10.03 m, respectively. The capacity of the lake is 14 million liters and the overflow feeds water to the Jaisamand lake. It serves for the agricultural, drinking and industrial demand of water of Alwar city and its adjoining villages. The open aqueduct from Siliserh lake, being constructed in 1845, delivers water to Alwar city and supports a population of about 3.15 lakh (2011 census) apart from delivering the overflow for irrigation in the downstream stretch. The lake also supports tourism, recreational activities, aquaculture (pisciculture and production of chestnut), agricultural production along its banks, conservation of biodiversity (wild crocodiles), apart from attracting migratory birds during winters. The employment generation, improved standard of living, and reduced migration of marginalized communities is the added socio-economic contribution of Siliserh lake. The soil around the lake is mainly alluvial and non-calcareous in texture; climate is classified as semi-arid to arid with average maximum and minimum ambient temperature of 38 ºC and 28 ºC, respectively. During the year 2018, average annual rainfall of about 350 mm spread over 80 rain days, and average sun hours of 9.5 per day with a total of 276 sun days were observed.

2.2 Sample collection and analysis

The present study was undertaken at Siliserh lake of Alwar city, Rajasthan. To determine water quality, sixteen (16) representative samples of surface water were collected (Fig. 1) covering entire water surface of the lake.

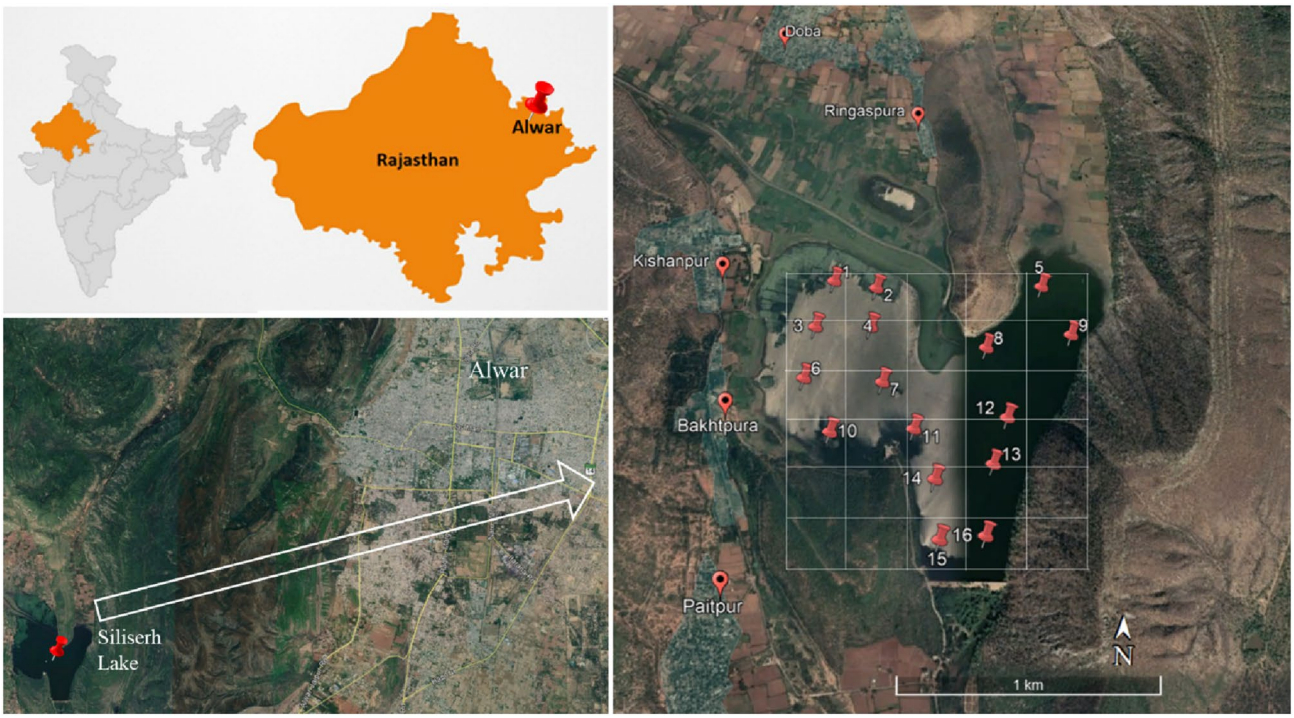


Fig. 1 Location of sampling stations in Siliserh Lake, Alwar, Rajasthan

The samples were collected in pre-cleaned and rinsed fresh poly-ethylene-terephthalate (PET) bottles which were filled upto the brim. The representative samples (04 in numbers) of sediments were also collected using a grab sampler. Some of the parameters of water (pH, electrical conductivity (EC), REDOX potential (ORP), Total dissolved solids (TDS), and dissolved oxygen (DO) were determined on-site using a multi parameter meter (Orion make, USA; Star A320 model). Other parameters such as total alkalinity (TA), total hardness (TH), chloride (Cl^-) were determined on-site using volumetric analyses. The samples were stored in an ice box at low temperature (4 ℃), transported to the laboratory and were analyzed within 24 h for rest of the parameters. The analyses for all the parameters were performed in triplicates (confirming the deviation on < 5%) using analytical grade (AR) chemicals and Type-1 ultrapure water as per the methods prescribed by APHA (2005). The estimation of phosphate (PO_4^{3-}), sulfate (SO_4^{2-}), nitrate (NO_3^-) and iron (Fe) was performed on a double beam UV–VIS spectrophotometer (LabIndia make UV3092 model). Sodium (Na^+), potassium (K^+), calcium (Ca^{2+}) and lithium (Li^+) were analyzed by flame photometry (Systronics make 128 µC model), and the minerals in sediments were determined by Fourier Transmission Infra-Red (FTIR) spectroscopy (Perkin Elmer make Spectrum 2 model). To characterize the quality of water for drinking, Water Quality Index (WQI) was computed based on the concentration of TDS, TH, TA, Ca^{2+}, Mg^{2+}, Fe, Cl^-, SO_4^{2-}, and NO_3^- (Haritash et al 2017). The equation used for calculation is as given as follows:

$$WQI = \Sigma WiQi/\Sigma Wi \quad (1)$$

where ***Qi*** is the quality rating for ith parameter [$Qi = 100 * (Vi - Vo)/(Si - Vo)$]; ***Vi*** is the obtained value; ***Vo*** is the ideal value in pure water (it is 0 for all parameters except pH—7.0); ***Si*** is the standard value; ***Wi*** is the weight of the parameter ($Wi = K/Si$); ***K*** is the proportionality constant ($K = 1/\Sigma(1/Si)$).

2.3 Statistical analysis

Statistical analysis of the results was done using Microsoft Excel (Ms-Excel) and Statistical Package for Social Scientists (SPSS) ver. 25. The sum of cations and anions (meq/L) was used to determine the ion balance using the equation $IB = \sum M^+/\sum M^-$; where M^+ is summation of cations (Na^+, K^+, Ca^{2+} and Mg^{2+}); M^- is summation of anions (CO_3^{2-}, HCO_3^-, Cl^-, SO_4, NO_3^-, PO_4^{3-} and SiO_4). The average value of IB was 0.97, and TDS/EC ratio of 0.5, which confirms the accuracy and acceptability of analysis (Chitrakshi and Haritash 2018). The matrix of correlation coefficient was generated to understand the inter-relations of different parameters. In order to arrive at the group of major parameters affecting water quality, principal component analysis (PCA) was performed to identify the possible sources. Finally, in order to classify the spatial variations and sources associated, cluster analysis was performed over SPSS 25.

3 Results and discussion

The results obtained upon physico-chemical characterization of water are given in Table 1. Assessment of suitability for drinking was done by comparing the observed values against the standards as prescribed by Bureau of Indian Standards (BIS) and World Health Organization (WHO, 2006). The average pH of the Siliserh Lake was 8.4 ranging from 7.3 to 9.5. The basic nature of pH shows dominance of alkaline species such as Ca^{2+} and Mg^{2+} over the acidic species such as Cl^- and SO_4^{2-}. The temperature of the lake water was in the range of 31–37 ℃ with an average value of 32 ℃, and average TDS was observed to be 112 mg/l which was much lower than the prescribed BIS standard of 500 mg/l for drinking. EC, which is the measurement of ionic strength of solution, varied between 152 and 248 µs/cm with the average value of 224 µS/cm. Secchi depth (SD) is the indirect measure of turbidity and eutrophication in water, and it varied between 0.40 and 1.1 m with an average value of 1.0 m. Since algal growth was limited, suspended impurities (average TSS value of 26 mg/l) contributed to turbid conditions. Spatial variations with respect to SD within the lake may be an outcome of variation in the growth of aquatic weeds, algae, and suspended solids.℃

3.1 Suitability for drinking

Total alkalinity (TA) of the samples of lake Siliserh varied between 64 and 96 mg/l with an average value of 84 mg/l. TA of all the water samples of Siliserh lake was found below the BIS standard of 200 mg/l. Based on TA, the concentration of carbonates and bicarbonates was also deduced. Since the average pH of water was 8.5, most of the alkalinity existed in the form of bicarbonates. Bicarbonates had an average value of 40 mg/l with the range extending from 10 to 58 mg/l. Some of the locations with relatively higher pH values reported the presence of carbonate ions. The average

Table 1 Physico-chemical characteristics of water of Siliserh lake during April, 2018

S. no	Temp (℃)	pH	EC (µS/cm)	TDS (mg/l)	TSS (mg/l)	SD (meter)	DO (mg/l)	TH (mg/l)	TA (mg/l)	Ca^{2+} (mg/l)	Mg^{2+} (mg/l)	Na^{+} (mg/l)	K^{+} (mg/l)	Li^{+} (mg/l)	Fe (mg/l)	CO_3^{2-} (mg/l)	HCO_3^{-} (mg/l)	Cl^{-} (mg/l)	SO_4^{2-} (mg/l)	SiO_4 (mg/l)	TP (mg/l)	NO_3^{-} (mg/l)	BOD_5 (mg/l)	NH_3 (mg/l)
1	32.8	8.9	209	104	44	0.5	7.2	124	80	7.2	0	5.16	2.79	0.02	0.16	10	29	8	13	29	1.5	17	4	1
2	32.8	7.3	248	124	70	0.55	7.3	228	96	9.0	36	5.89	5.71	0.04	0.16	0	58	12	6	9.0	1.0	8.8	4	0.9
3	32.9	9.5	154	77	48	0.40	7.3	104	64	6.6	0	5.87	4.57	0.04	0.19	14	10	20	7	7.6	1.7	22	3	1.2
4	33.2	8.6	245	123	32	0.65	7.5	172	96	8.5	21	5.76	5.83	0.04	0.04	0	58	16	5	8.4	0.8	5	3	0.8
5	33.1	8.1	245	122	20	0.55	7.5	148	80	8.7	12	5.91	5.66	0.04	0.10	0	48	24	4	8.4	0.9	6	3	0.4
6	33.2	8.9	185	92	34	1.10	7.4	152	64	7.2	19	5.72	5.46	0.04	0.03	14	10	16	4	7.4	0.7	16	3	1.4
7	30.7	9.0	200	100	16	0.66	7.2	76	80	8.0	1	5.72	5.56	0.05	0.04	10	29	12	6	8.4	1.0	10	4	1.4
8	30.7	8.5	242	122	20	0.65	7.2	112	96	9.0	9	6.02	5.40	0.04	0.03	0	58	16	5	8.8	0.9	8	4	0.8
9	36.9	8.4	239	120	20	0.56	7.6	124	80	9.1	11	5.07	5.87	0.04	0.07	0	48	20	8	8.0	0.8	8	4	0.5
10	31.1	8.0	224	112	34	1.10	7.2	136	88	8.5	13	5.60	5.71	0.04	0.04	5	44	20	3	7.3	1.0	11	4	0.5
11	31.4	8.6	228	114	12	0.81	7.1	220	72	9.0	34	5.87	5.79	0.05	0.10	10	24	12	5	7.5	0.9	12	4	1.2
12	33.7	7.8	226	113	12	0.61	7.5	132	88	9.1	15	5.78	5.62	0.05	0.11	10	34	20	5	7.7	1.2	11	4	0.9
13	31.8	8.1	236	118	8	0.58	8.0	136	80	8.9	10	5.79	5.80	0.05	0.08	0	48	20	5	7.9	1.4	9	3	0.8
14	31.7	8.8	236	118	16	0.71	9.1	116	96	9.3	10	5.60	5.97	0.05	0.07	0	58	20	5	9.4	1.0	10	2	0.5
15	31.7	8.8	234	117	16	0.99	8.6	116	88	9.1	13	5.66	5.83	0.05	0.12	10	34	12	6	7.8	1.1	12	2	0.9
16	31.8	8.5	236	118	20	0.96	8.3	104	96	9.3	0	5.76	5.89	0.04	0.07	5	48	24	5	7.7	1.5	11	2	1.1
Range	30.7–36.9	7.3—9.5	154–248	77–124	48–70	0.4–1.1	7.3–9.1	76–228	64–96	7.2–9.3	0–36	5.07–6.02	2.79–5.97	0.02–0.05	0.03–0.19	0–14	10–58	8–24	3–13	7.3–29	0.7–1.5	5–17	2–4	0.4–1.4
Mean	32	8.5	224	112	26	0.71	7.6	138	84	8.5	12.75	5.70	5.47	0.04	0.09	5.48	39.88	17.00	5.75	9.40	1.09	11.08	3.0	1.0
SD(±)	1.5	0.5	25.5	12.9	16.5	0.2	0.6	40.4	10.9	0.8	10.8	0.3	0.8	0.01	0.05	5.53	16.20	4.73	2.27	5.29	0.29	4.23	0.79	0.32

concentration of carbonates in lake water was 5.5 mg/l. Spatial variations of carbonate may also be attributed to dissolution of carbonate minerals from the soil/rocks present in the region which had resulted in traceable concentration of carbonates at some locations, particularly along the western bank of the lake. The other reason could be addition of domestic wastewater which has presence of carbonates since carbonates are components of detergents. The chemical species such as HCO_3^-, CO_3^{2-}, Ca, and Mg are also responsible for imparting hardness in water. Total hardness (as $CaCO_3$) was reported to vary between 76 and 228 mg/l with an average value of 138 mg/l. Based on TH, most of the samples were classified as soft to moderately hard.

Based on the observed values of cations in the water, Mg^{2+} was the most abundant species followed by calcium, sodium, potassium and lithium. It was observed that magnesium had an average value of 13 mg/l ranging from the below detectable limit (BDL) to 36 mg/l. Comparing it with the standard prescribed value of 30 mg/l for drinking water as per BIS standard, it was found that 87% of the samples were within safe limit, and only two samples exceeded the prescribed standard. Dissolved concentration of calcium regulates hardness of water, and so it is an important chemical species to determine suitability of drinking water. Ca varied between 7.2 and 9.3 mg/l with a mean value of 8.5 mg/l, much lower than the prescribed standard of 75 mg/l. Hence, none of the sample exceeded permissible limit of calcium in the study area. Similarly, concentration of sodium was also found to be low ranging from 5.07 to 6.02 mg/l with an average value of 5.70 mg/l. Based on relative concentration of cations, alkaline earth metals (Ca^{2+} and Mg^{2+}) fairly exceeded the alkali metals (Na^+ and K^+). Dominance of earth metals is responsible for the basic pH of water throughout the region. On the other hand, lower levels of Na and K resulted in relatively lower concentration of salinity. This indicated that the water is alkaline but non-saline, which is an indication of good quality of water. Moreover, the levels of studied cations were well within the limits prescribed for drinking water pointing towards absence of a major source of pollution. Significant positive correlation of Ca with Li (which is purely crustal in origin) strengthens the assumption that dominant fraction of cations comes from natural/crustal source. Similarly, traceable concentration of dissolved iron (Fe) in water with an average value of 0.09 mg/l confirmed the absence of anthropogenic addition of cations.

Among the anions, the concentration of chloride ions varied from 8 to 24 mg/l with an average value of 17 mg/l

Table 2 Percent suitability and water quality index of Siliserh lake for drinking as against the prescribed standards

S. no	Parameter (unit)	Prescribed limit		Mean; range (Min.–Max.)	Suitable samples (%)	
		WHO	BIS		WHO	BIS
1	pH	6.5–8.5	6.5–8.5	8.49 (7.3–9.5)	50	50
2	EC (µS/cm)	-NM-	-NM-	224 (154–248)	-NM-	-NM-
3	TDS (mg/l)	500	500	142 (77–124)	100	100
5	TH (mg/l)	500	200	138 (76–228)	100	93
6	TA (mg/l)	250	200	84 (64–96)	100	100
9	Na^+ (mg/l)	-NM-	-NM-	5.7 (5.07–6.02)	-NM-	-NM-
10	K^+ (mg/l)	-NM-	-NM-	5.4 (2.8–5.9)	-NM-	-NM-
7	Ca^2+ (mg/l)	-NM-	75	8.5 (6.6–9.3)	-NM-	100
8	Mg^{2+} (mg/l)	50	30	13 (0–36)	100	87
11	Li^{2+} (mg/l)	-NM-	-NM-	0.04 (0.02–0.05)	-NM-	-NM-
4	Fe (mg/l)	0.3	0.3	0.09 (0.03–0.19)	100	100
12	CO_3^{2-} (mg/l)	-NM-	-NM-	6 (0–14)	-NM-	-NM-
13	HCO_3^- (mg/l)	-NM-	-NM-	40 (10–58)	-NM-	-NM-
14	Cl^- (mg/l)	200	250	17 (8–24)	100	100
15	SO_4^{2-} (mg/l)	250	200	6 (3–13)	100	100
16	NO_3^- (mg/l)	50	45	11 (5.3–17)	100	100
17	PO_4^2- (mg/l)	-NM-	-NM-	1.1 (0.7–1.7)	-NM-	-NM-
18	WQI*			21–78	–	75

which is significantly lower than the prescribed standard for chloride, i.e., 250 mg/l (BIS 2012). Similarly, low concentration of sulfate throughout the lake confirmed absence of addition of sewage or gray water. The SO_4^{2-} varied from 3 to 13 mg/l with an average value of 5.8 mg/l, and it is well within the permissible limit specified for sulfate (200 mg/l) as per BIS (2012). Significant positive correlation was observed for concentration of sulfate with dissolved silica (which is purely crustal in origin) confirming that sulfate comes from crustal source. Based on relative abundance, average concentration of bicarbonate ions was maximum (40 mg/l), followed by chloride (17 mg/l), silicate (9.4 mg/l), sulfate (5.8 mg/l), and carbonate (5.5 mg/l). The sum of carbonates and bicarbonates ($CO_3^{2-} + HCO_3^-$) was almost double than the sum of chlorides and sulfate ($Cl^- + SO_4^{2-}$), which indicates that the basic anions dominate the acidic anions adding to basic nature of water. In the present study, most of the samples were found suitable for drinking and on the basis of WQI it was observed that about 75% of samples (Table 2) collected from the lake are suitable for drinking, but after preliminary treatment and disinfection.

3.2 Suitability for aquatic life

An important parameter to establish water quality for aquatic life is oxidizing/reducing environment present in the lake ecosystem. Based on the values of ORP, the conditions were found to be oxidizing in nature. Although there were some sources of organic impurities in water, still the conditions were oxidizing with an average value of 82 mV for ORP. The minimum ORP (50 mV) was observed at the location 1 (near Kishanpura village) (Fig. 1) and maximum (112 mV) at location 3 (near Ringaspura village) where the depth was shallow and rich growth of algae and other aquatic plants was observed. Since submerged aquatic vegetation and algae carries out photosynthesis, and oxygen is the byproduct of this process which eventually adds DO to surface water. This could be reason for elevated ORP level at location 3. The relatively high levels of DO throughout the lake (average concentration of 8 mg/l) confirmed the prevalence of oxidizing conditions. Biochemical oxygen demand (BOD_5) is another key parameter to establish biological quality of water, and it was in the range of 2–4 mg/l with an average value of 3 mg/l. The monsoonal rains are major source of water to lake Siliserh with an average annual rainfall of 675 mm. Flushing of organic matter from the catchment area of about 11 km^2 results in addition of organic load. For rest of the period, addition of organic matter takes place from within the lake due to death and decomposition of microbial and plant biomass. This could be the probable reason for relatively lower values of BOD_5. Lower levels of organic impurities result in sufficiently good level of dissolved oxygen which supports aquatic life and oxidizes impurities. Nitrogen is the other important parameter and it is estimated as dissolved ammonia and nitrate in water. Since the BOD_5 values were low, the fraction of organic nitrogen in water is considered to be low. Had there been presence of organic nitrogen, relatively high values of BOD_5 would have been obtained. The ammonical fraction (dissolved ammonia) of nitrogen varied from 0.4 mg/l to 1.4 mg/l with an average value of 1.0 mg/l. The prevalent levels of dissolved ammonia are considered to be safe for aquatic life as compared against the limit of 1.4 mg/l as per CCME 2007. Phosphate is a critical parameter with respect to the eutrophication in waterbody. Excess of nutrients is the reason for high biological activity of primary producers such as phytoplankton and algae. In this study, the average value of phosphate was found to be 1.09 mg/l varying between 0.7 and 1.7 mg/l. Maximum value was observed at location 3 and minimum at location 6. With respect to the concentration of phosphate, lake Siliserh was found to be eutrophic in nature. Nitrate was observed to be in the range of 5.3 to 17 mg/l. Minimum value (5.3 mg/l) was observed at sampling location 4, and maximum value (17 mg/l) at sampling location 1 near the village Ringaspura. Slightly raised level of nitrate at Ringaspura may be because of addition of domestic waste water from the adjoining villages and surface runoff from the agricultural fields since nitrate is the oxidized form of nitrogen, a major component of fertilizer used for crops. The high concentration of nitrate is known to support the growth of algae. Based on the analysis of nitrogen and phosphorus, it was seen that the lake is near-eutrophic and had the N:P ratio of about 10:1 which may be categorized as transitional phase towards eutrophication (Yun et al. 2016). Despite being classified as eutrophic based on concentration of phosphate and nitrate, limited growth of algae was observed in the lake which may be attributed to a functional food chain in which the nutrients are transferred from on level to another and finally being exported with the fish catch (Singh & Haritash 2019; Pipil et al. 2021).

3.3 Suitability for irrigation

The parameters important for determining the quality for agricultural use are dominantly affected by Sodium (Na). The Percent sodium (%Na) ($\%\ Na = [(Na^+ + K^+) \times 100]/(Ca^{2+} + Mg^{2+} + K^+)$) (Wilcox, 1955) is an important parameter to determine sodium hazard since high percentage of sodium in irrigation water restricts plant growth. It reacts

Table 3 Percent suitability of water of Siliserh Lake for irrigation against various indices

S. no	Parameter	Mean value (Range: min–max)	Category	Inference	Suitable samples (%)
1	EC	224 (154–248)	250[a]	Very good	100
			250–750	Good	0
			750–2000	Marginal	0
			2000–3000	Very poor	0
			3000	Harmful	0
2	Na (%)	26 (10–53)	20[a]	Very good	31
			20–40	Good	44
			40–60	Marginal	25
			60–80	Poor	0
			80	Harmful	0
3	RSC	− 0.7 (− 2.6–0.5)	< 0[b]	Very good	75
			0–2.5	Good	25
			2.5–5.0	Marginal	0
			5.0–7.5	Poor	0
			> 7.5	Harmful	0
4	SAR	0.34 (0.19–0.63)	0-10[c]	Very good	100
			10–18	Good	0
			18–26	Marginal-poor	0
			> 26	Harmful	0
5	Mg Hazard	56 (0–87)	< 50%[d]	Good	25
			> 50%	Harmful	75
7	KR	0.29 (0.07–0.77)	< 1[e]	Good–very good	100
			1–2	Marginal	0
			> 2	Poor-harmful	0
8	PI	76.72 (24.85–158.55)	Class I	Safe	6
			Class II	Moderate	63
			Class III	Doubtful	31

[a]Wilcox (1955)
[b]Bishnoi et al. (1984)
[c]Richards (1954)
[d]Paliwal (1972)
[e]Kelly (1963)

with the soil and reduces its permeability. Finer soil texture with high organic matter content has more impact of sodium (Haritash et al. 2008). The classification (Wilcox 1995) for different ratios is given in Table 3. In the present study, the average value of percent sodium was found to be 26%, which shows that the water of Siliserh lake is good for irrigation. Similarly, soluble sodium percentage (SSP) (Eaton 1950) ranged from 7 to 44% with an average value of 19%. Based on the classification, the surface water of Siliserh lake was classified as good for irrigation. Another important parameter, Residual sodium carbonate (RSC) determines the excess sodium carbonate ($RSC = (HCO_3^- + CO_3^{2-}) - (Ca^{2+} + Mg^{2+})$) which remains after calcium and magnesium have been precipitated with carbonate (Eaton 1950). Excess sodium carbonate reduces the permeability of soil as excess of RSC adsorb over the soil particles affecting the physical properties of soil (Aggarwal et al., 2021). The average value of RSC is − 0.7 in the present study. Negative value of RSC indicates that the water is suitable for irrigation. SAR ($SAR = Na + /(Ca^{2+} + Mg^{2+}/2)^{1/2}$) is another key

parameter to determine the suitability of water for irrigation, and it represents the alkali and sodium hazard. SAR in the present study indicated very good quality with all the samples lying in low salinity and low sodium hazard.

The Kelly's ratio (KR) is used to access the sodicity hazard ($KR = Na + /(Ca^{2+} + Mg^{2+})$) associated with water to be used for irrigation of water (Kelly 1963). KR ranged from 0.07 to 0.77 with a mean value of 0.27, which is within the permissible limit (< 1.0). This indicates that water can be used over long period of time for irrigation purpose. The Magnesium hazard (Mg Hz) ($Mg\ Hazard = (Mg^{2+} \times 100)/(Ca^{2+} + Mg^{2+})$) is another parameter for determining the suitability of water for irrigation (Paliwal 1972) and its high concentration in water may disrupt the cation exchange procedure of soil particles. In the present study, average value of magnesium was found to be 56% indicating that the water quality is compromised due to high concentration of dissolved magnesium. Long-term irrigation of soil is influenced by Na, Ca, Mg, and HCO_3^- content of the water affects soil permeability. The Permeability index (PI) (Doneen 1964) ($PI = [Na^+ + (HCO_3^-)^{1/2}/(Ca^{2+} + Mg^{2+} + Na^+)] \times 100$) is another ratio to calculate suitability of water for irrigation. Based on it, 70% of the samples were found fit for irrigation and 30% were found unfit for irrigation. Since the agricultural soil around Siliserh lake is sandy and permeable, it results in no accumulation of Na in soil, thereby making the water from Siliserh lake fit for irrigation.

3.4 Suitability for industrial use

Water for industrial use involves applications such as boiler feedwater, cooling water, washing, and other manufacturing processes. In order to prevent corrosion and scaling in utilities, water quality must be monitored from time to time. In this study, to determine the extent of corrosion in the pipes and in utilities due to dissolution of carbonates, saturation indices were calculated. The Langelier Saturation Index (LSI) and Ryznar Stability Index (RSI) are used to determine the stability of calcium carbonate in water (Langelier 1936). It indicates whether the calcium carbonate dissolved in water will precipitate or be in equilibrium with precipitated $CaCO_3$. Since the values of LSI are negative, the analyzed samples of water had limited potential towards scaling (only 6% samples). Hence, water of Siliserh lake was found to be safe against the problem of scaling. Ryznar stability index (RSI) is used to determine aggressiveness (Ryznar 1944) and corrosion potential (Carrier 1965) of the water. As per the Carrier (1965), RSI values indicate that 96% of samples (RSI > 9) had intolerable corrosion potential while remaining 6% had heavy corrosion potential in industrial utilities. As per Ryznar**,** all the samples were classified as aggressive, thus unsuitable for transportation through pipes. Similarly, water with the Corrosivity Ratio (CR) less than unity is considered to be safe for transport of water through pipes of any material and the corrosivity ratio more than unity indicates corrosive nature and hence cannot be transported through

Table 4 Appraisal of water of Siliserh Lake for industrial use

S. no	Parameter	Mean value (Range: min–max)	Classification		Samples (%)
			Value	Inference	
1	LSI	− 0.55 (− 1.5–0.012)	< 0.0	No scaling	94%
			> 0.0	Scaling	6%
2	RSI* (Carrier, 1965)	9.54 (8.8–10)	4–5	Heavy scale	-Nil-
			5–6	Light scale	-Nil-
			6–7	Little scale/corrosion	-Nil-
			7–7.5	Corrosion significant	-Nil-
			7.5–9	Heavy corrosion	6%
			> 9	Corrosion intolerable	94%
3	CR	0.73 (0.49–1.12)	< 1	Safe	94%
			> 1	Not safe	6%
4	CC (Wu et al., 2015)	0.41 (− 0.79–2.46)	> 0	Corrosive	63%
			< 0	Not corrosive	37%
5	FC (Wu et al., 2015)	29.30 (22.19–43.50)	< 60	No foaming action	100%
			60–200	Marginal foaming	-Nil-
			> 200	Foaming action	- Nil-

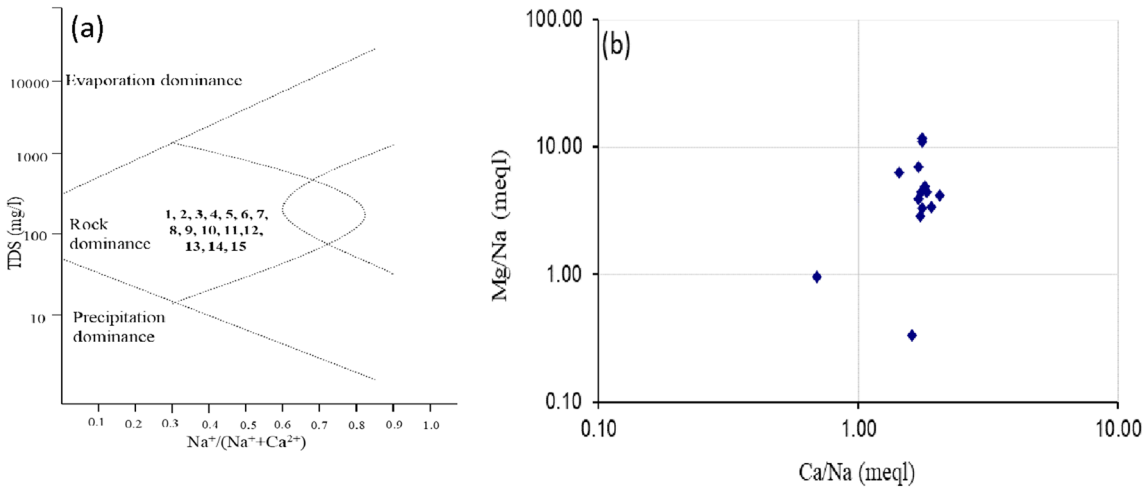


Fig. 2 Gibbs boomerang plot (**a**) for mechanism governing water chemistry and **b** weathering type based on Na-normalized Ca and Mg

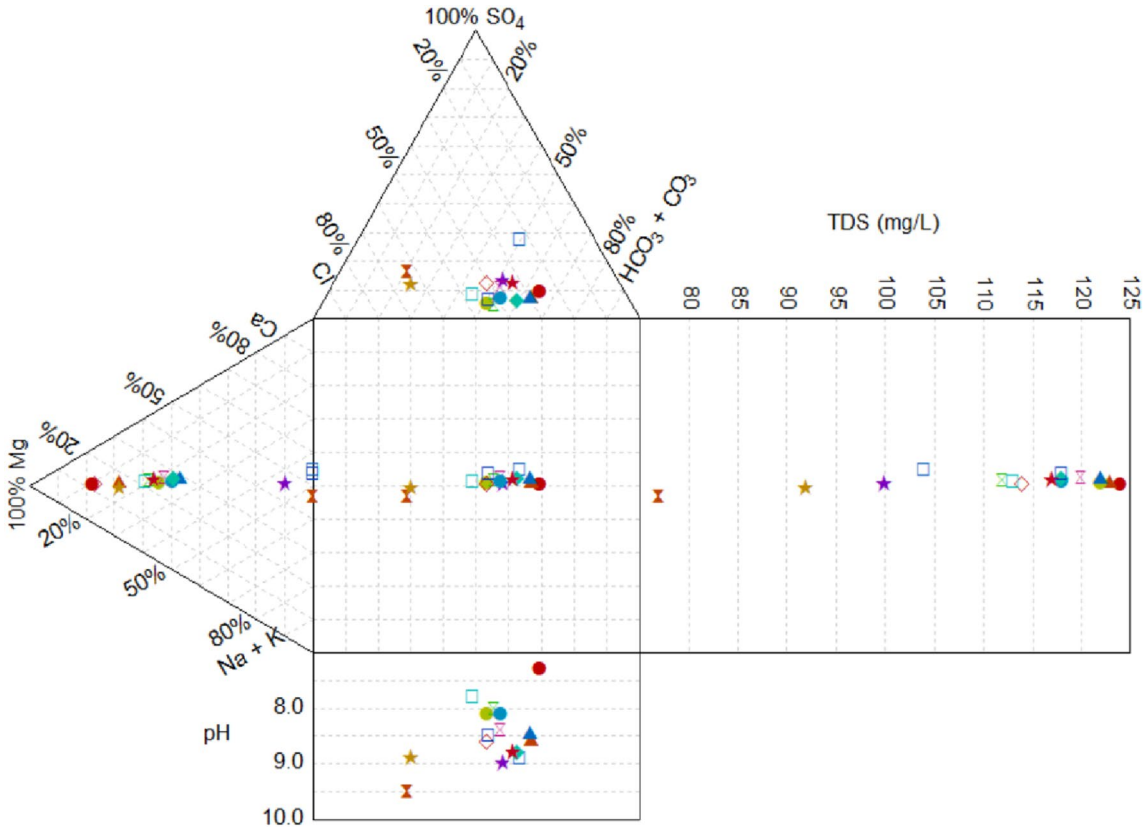


Fig. 3 Durov's diagram for geochemical classification and chemical type of water of Siliserh lake

metal pipes (Ryner 1944; Raman 1985). In the present study, average value of corrosivity ratio was found to be 0.82 varying from 0.49 to 1.12. Values for almost all the samples were below unity indicating that most of the samples had limited corrosion potential. The Corrosion Coefficient (CC) is another indicator that reflects the intensity of corrosion. Positive values of corrosion coefficient indicate that water is corrosive while negative value shows non-corrosive nature of water. As per the data (Table 4), 63% water samples of the Siliserh lake were found to be corrosive in nature while remaining 37% of water samples were found to be non-corrosive in nature. Foaming is another important parameter for determining the use of water in industry. It is the continuous formation of the bubbles when water is boiled. Bubbles and froth build up on the surface of the boiler and elevate the water level in boilers resulting hindrance in its working. Values of foaming coefficient (FC) less than 60 lead to no foaming action whereas, values of foaming above 200 lead to foaming action. As per the obtained values of foaming coefficient for the Siliserh lake water, all the values were below 60. Hence, Siliserh lake water is safe against the action of foaming.

3.5 Geochemical classification and weathering profile

The chemical interaction of water with surrounding material/rocks is the dominant route for dissolution of ions apart from atmospheric deposition (Subramanian 1987). The dominant route for addition of cations and anions is generally ascertained by Gibbs boomerang plot (Gibbs 1970). Gibb's classification (Fig. 2a) revealed that rock–water interaction is the process regulating chemistry of water in Siliserh lake. Further, the type (silicate/carbonate) of weathering can be determined calculating the relative ratio of cations and anions. Generally, silicate weathering represents Ca/Na ratio of 0.35 ± 0.15, HCO_3/Na 2 ± 1, and Mg/Na 0.24 ± 0.12 (Gaillardet et al. 1999); whereas carbonate weathering represents Ca/Na of 50, HCO3/Na of 120, and Mg/Na of 10 (Meybeck 1986). In the present study, silicate weathering was observed since average Ca/Na ratio was 1.5, HCO_3/Na ratio close to 6.99 and Mg/Na ratio of 2.23 which is significantly lower than the aquifers regulated by carbonate geology. In addition to this, the sodium normalized calcium and magnesium (Fig. 2b) confirmed that silicate weathering is regulating the water chemistry of Siliserh lake. The value of (Ca + Mg)/(Na + K), and (Na + K)/TZ^+ was close to 1 and 0.3, respectively, confirming that silicate weathering is the dominant process. The relative dominance of different cations and anions can be represented using Durov's diagram (Durov 1948) which can help to classify the dominant chemical composition/type of water. Durov's classification (Fig. 3) revealed that the water of Siliserh lake was dominantly $MgHCO_3$ type followed by CaMgCl type. Based on relative abundance of ions, the alkaline earth metal (Ca + Mg) exceeded the alkaline metals (Na + K); whereas strong acids (Cl^- and SO_4^{2-}) overtook weak acids (HCO_3^- and CO_3^{2-}).

3.6 Analysis of sediments

Sediments in a waterbody act as carriers as well as sinks for the contaminants, and sediments of a lake are replica of long-term course of contamination or anthropogenic input (Benamar et al. 2019). The concentration of total organic carbon (TOC) and orthophosphate, and their mineralogical composition is given in Table 5. It was observed that a significant fraction of the sediments was composed of potassium, calcium, sodium, and lithium which are a part of the rock material (crust) present in the region. The sediments had fairly good organic carbon (TOC) associated with it ranging from 3.1 to 18.7% with an average value of 8.1%. Relatively high concentration of sediment-associated organic matter was observed near the dam (18.7) since most of the suspended organic fraction settles to the bottom in this region. Similar to organic carbon, relatively high concentration of phosphate was observed near the dam. An almost equal average concentration of phosphate in sediments and water suggests of near equilibrium conditions *w.r.t* partitioning of phosphate, indicating that the sediments are almost at equilibrium with the mobile phase, i.e., water, in respect of orthophosphate and further increase in phosphate in water may not result in binding with sediments, thus strengthening the possibility of eutrophication. FTIR

Table 5 Chemical and mineralogical composition of bottom sediments of Siliserh lake

S. no	TOC (%)	PO_4^{3-} (mg/g)	Chief minerals	Chemical formula
1	4.3	0.69	Kaolinite	$Al_2\ Si_2O_5(OH)_4$
2	3.1	0.65	Gypsum	$CaSO_4.2H_2O$
3	18.7	0.68	Calcite	$CaCO_3$
4	6.2	0.93	Hemihydrate of CaSO4	$CaSO_4.0.5H_2O$

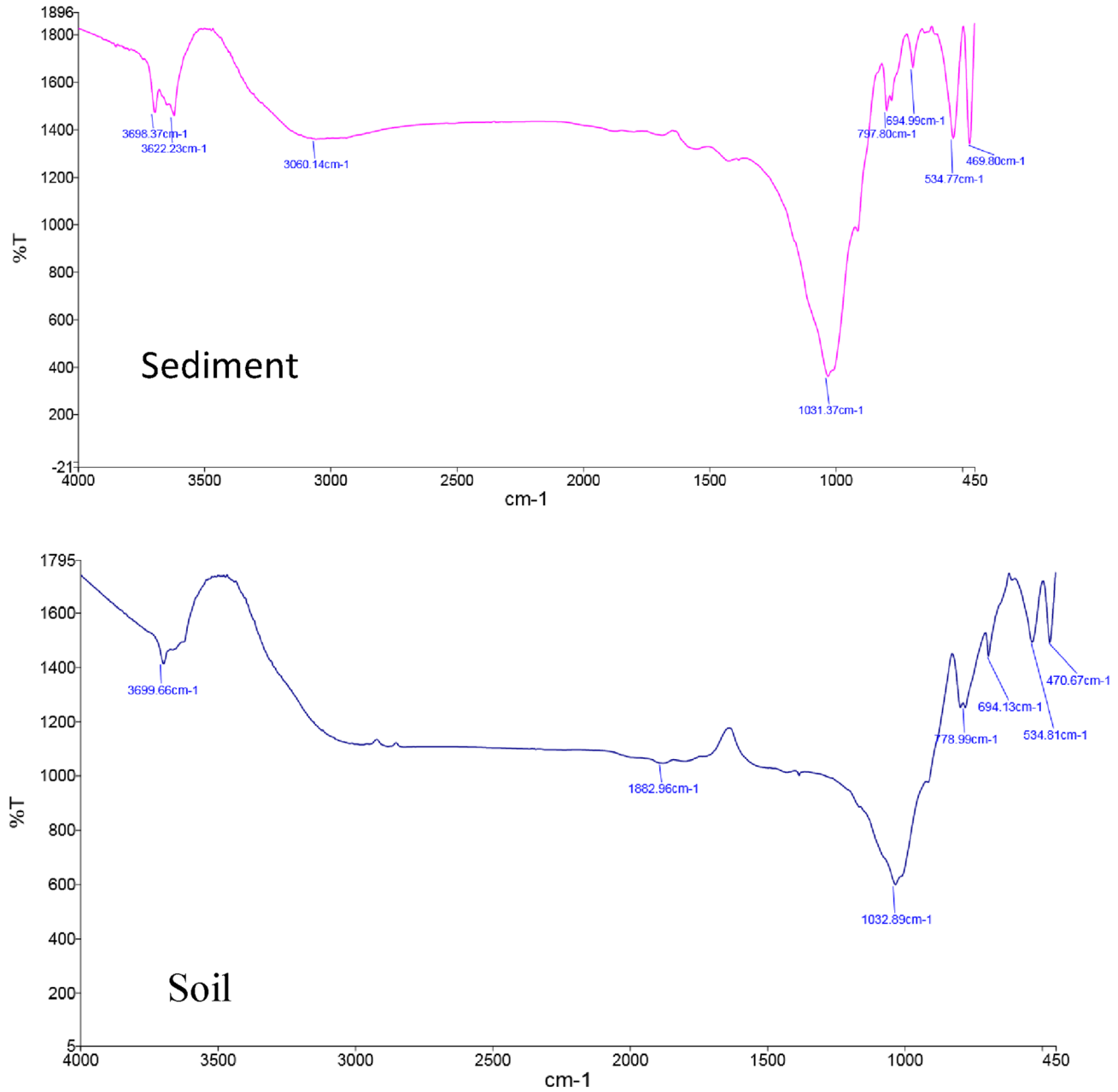


Fig. 4 FTIR spectra of sediment and soil in and around Siliserh lake

analysis of the sediments revealed that Kaolinite clay was the major mineral present in the sediments, and it was followed by the presence of gypsum, hemihydrate of gypsum, and calcite (Table 5; Fig. 4). A correlation was observed between the chemical composition of dominant minerals and physico-chemical characteristics of water in Siliserh Lake, thus confirming that rock–water interaction is dominantly regulating the water quality. Since the mineralogical composition of soil around Siliserh lake and its sediments is same, i.e., it is chiefly composed of kaolinite; therefore, it is assumed that most of the addition of suspended solids as well as dissolved nutrients through runoff is from nearby region, and long-range transport of sediments and pollutants into Siliserh lake may be ruled out. Moreover, presence of kaolinite as end-member of

geochemical breakdown confirms that silicate weathering is dominantly operating in the region around Siliserh lake.

3.7 Statistical analysis

The correlation matrix revealed that there is a positive correlation between the alkalinity and HCO_3^- ions; whereas the pH was positively related to CO_3^{2-}. Similarly, Na, K, and Li were observed to strongly related to each other indicating a common source. The nitrate, phosphate, and ammonia were also observed to be positively related. The results were also subjected to Principal Component Analysis for a better understanding about the source (Table 6). Component-1 revealed significant positive correlation among EC, Alkalinity, Ca, and HCO_3^- which may be classified as crustal in origin. The hierarchical cluster analysis is formed clusters on the basis of spatial variations of analyzed parameters (Fig. 5). The sampling location 1, 2, and 3 had proximity to agricultural fields and, therefore, had different quality as compared to the cluster of location 4 to 16 which had low to negligible influence of agricultural activity.

Table 6 Rotated component matrix representing principal component analysis (PCA) for Siliserh lake, Alwar

Variables	Components				
	PC 1	*PC 2*	*PC 3*	*PC 4*	*PC 5*
pH	−0.662	−0.187	−0.464	−0.122	0.299
EC	0.903	0.055	0.277	−0.136	0.085
DO	0.291	0.106	−0.170	0.049	0.917
ORP	0.495	0.293	0.564	−0.362	−0.002
TH	0.105	0.051	0.944	0.062	−0.185
TA	0.863	0.078	−0.069	0.071	0.101
BOD	−0.014	−0.205	0.169	−0.098	−0.876
Na	−0.007	0.809	0.111	0.295	−0.195
K	0.350	0.760	0.205	−0.372	0.267
Ca	0.788	0.336	0.153	−0.087	0.243
Mg	0.114	0.239	0.927	−0.112	−0.123
Li	0.023	0.795	0.055	−0.117	0.309
Fe	−0.221	−0.328	0.270	0.798	0.086
CO_3^{2-}	−0.903	−0.026	−0.071	0.152	−0.015
HCO_3^-	0.971	0.045	0.022	−0.072	0.047
SO_4^{2-}	−0.143	−0.873	−0.122	0.284	−0.040
NO3	−0.819	−0.243	−0.139	0.399	0.074
PO_4^{3-}	−0.182	−0.159	−0.419	0.821	0.113
SiO_4^{4-}	−0.033	−0.885	−0.112	0.251	−0.134
NH_3	−0.746	0.146	−0.042	0.146	−0.152
Eigen values	8.130	3.450	2.990	1.570	1.050
CV%	40.630	57.860	72.800	80.660	85.890

4 Conclusion

Based on the physico-chemical characterization of water, the alkaline earth cations (Ca and Mg) exceeded alkali metals (Na and K), and relative abundance of cations and anions followed Mg > Ca > Na > K, and HCO3 > Cl > SO_4 > CO_3 trend. The water was classified as $MgHCO_3$ type and CaMgCl type based on the relative abundance of ions, and silicate weathering played a major role in genesis of water quality. The geochemical classification of water indicates that direct cation exchange results in addition of calcium and magnesium from the interacting rocks to water resulting in relatively higher concentration of these cations. The chemical constitution of the identified minerals correlated with the chemical composition of water confirming to natural source for most of the dissolved ions. Similarly, natural flushing of organic impurities during the rainfall is the major source for organic impurities. The analysis of sediments too confirmed the accumulation of organic matter and its autochthonous circulation, subsequently. The dominance of chloride among anions confirmed addition of domestic wastewater from the surrounding villages. Similarly, the N:P ratio value (10:1) representing gradual movement of Siliserh lake from oligotrophic to eutrophic status indicating that anthropogenic activities around the lake are leading to cultural eutrophication. The agricultural practices over the exposed land of Siliserh lake during dry season led to solubilization of nutrients and pesticide residues and suspension of loose soil during precipitation and runoff leading to accumulation of nutrients and contaminated sediments in the lake. The agrochemicals from the fields are inducing toxicity to water which in turn affect the exposed population and organisms such as zooplankton and fish. Since zooplankton and small fishes are the keystone species of an aquatic ecosystem, the active biological food chain may break leading to poor ecosystem function. Moreover, the higher forms such as waterfowls, migratory birds and endemic populations of crocodiles are at risk of increased toxicity and declining water quality. The dam over the lake accumulates sediments and organic carbon in the reservoir and enhanced rate of erosion and washing can reduce capacity/volume of reservoir and concentration of pollutants resulting in fast decline of water quality. Since most of the organic and inorganic pollutants in Siliserh lake are transported from the short distances within a closed catchment, it is easy to implement an intervention to maintain/improve the water quality. The study stresses that the early intervention can help to maintain the suitability of water for drinking, irrigation and aquatic life in Siliserh lake.

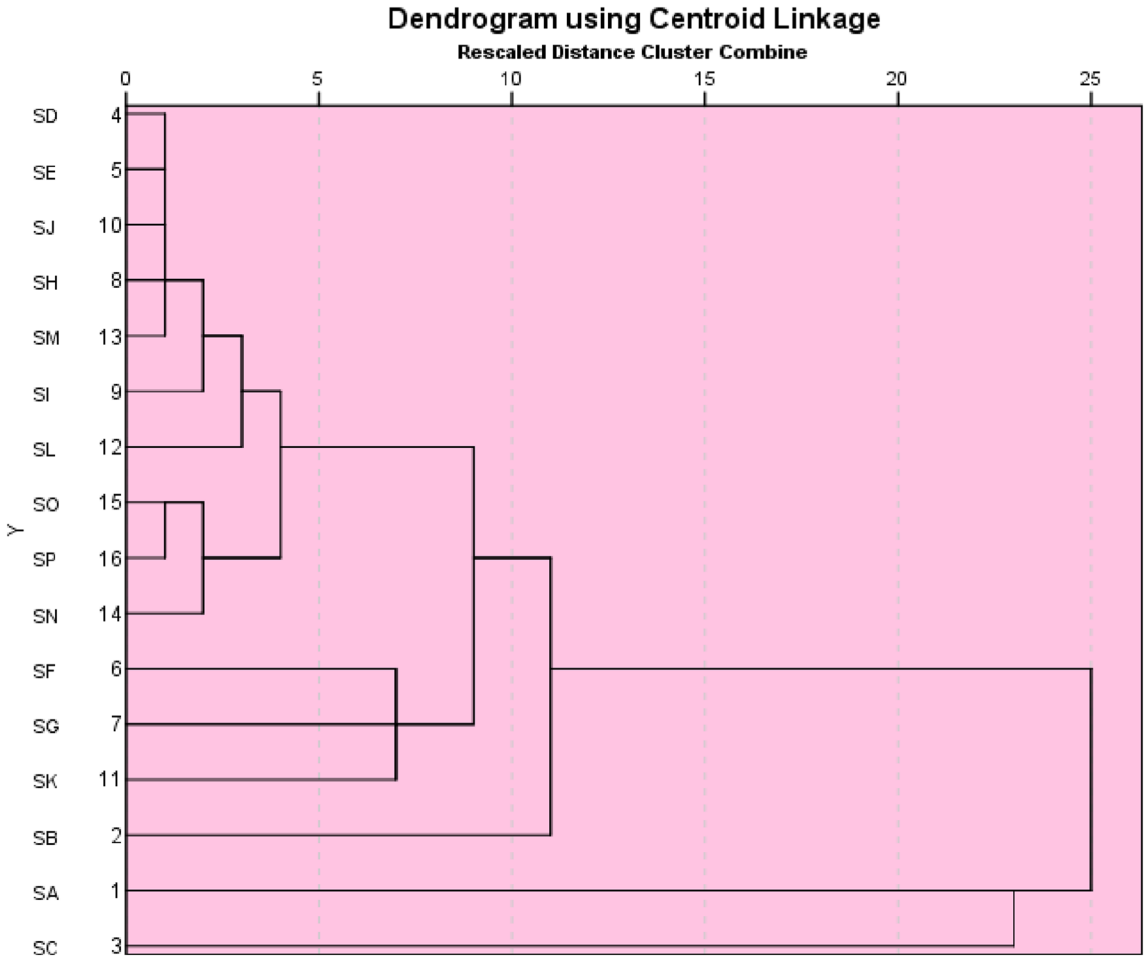


Fig. 5 Dendrogram for cluster analysis of water samples of Siliserh lake

Acknowledgement Authors acknowledge the help of Department of irrigation, Alwar, Rajasthan to grant permission for collection of samples; and Dr. Ram Singh, Department of Applied Chemistry, DTU for FTIR analysis

Author contributions The study was jointly designed and performed by Chitrakshi and A.K. Haritash

Funding No funding was received for this study

Data availability All data generated or analyzed during this study are included in this published article

Code availability Not applicable.

Declarations

Conflict of interests The authors declare that they have no conflict of interests

Ethics approval and consent to participate Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Aggarwal A, Soni J, Sharma K, Sapra M, Chitrakshi KO, Haritash AK (2021) Hydrogeochemical assessment of groundwater for drinking and agricultural use: a case study of rural areas of Alwar, Rajasthan. Environ Manage 67:513–552. https://doi.org/10.1007/s00267-020-01361-x

Ambastha SK, Haritash AK (2021) Prevalence and risk analysis of fluoride in groundwater around sandstone mine in Haryana, India. Rend Fis Acc Lincei 32:577–584. https://doi.org/10.1007/s12210-021-00997-z

American Public Health Association (APHA) (2005) Standard methods for the examination of water and wastewater, 21st edn. American Public Health Association, Washington DC

Banerjee RC, Upadhyay DS (1976) A survey of Drought and scarcity in Rajasthan. Proc Indian Natn Sci Acad 42(A1):15–21

Bayram A, Onsoy H (2015) Sand and gravel mining impact on the surface water quality: a case study from the city of Tirebolu (Giresun Province, NE Turkey). Environment Earth Science 73(5):1997–2011. https://doi.org/10.1007/s12665-014-3549-2

Benamar A, Tian Y, Portet-Koltalo F, Ammami MT, Giusti-Petrucciani N, Song Y, Boulange-Lecomte C (2019) Enhanced electrokinetic remediation of multi-contaminated dredged sediments and induced effect on their toxicity. Chemosphere 228:744–755. https://doi.org/10.1016/j.chemosphere.2019.04.063

BIS (2012) Indian standard drinking water specifications. Second Revision, Bureau of Indian Standards, New Delhi

Canadian Council of Minister of the Environment (CCME) (2007) For the protection of aquatic life (2007) In: Canadian environmental quality guideline, 1999. Canadian Council of Minister of the Environment, Winnipeg

Carrier air conditioning company (1965) Handbook of air conditioning system design. McGrawHill Books, New York

CGWB (2016) Ground water year book (2014–2015) Rajasthan state. Central Ground Water Board Regional office data centre, Western Region, Jaipur

Chitrakshi and Haritash AK, (2018) Hydrogeochemical characterization and suitability appraisal of groundwater around stone quarries in Mahendragarh. India Environmental Earth Sci 77:252. https://doi.org/10.1007/s12665-018-7431-5

De Donno A, Bagordo F, Serio F, Grassi T, Devoti G, Guido M (2017) Environmental quality and hygienic safety of the Alimini Lakes (Puglia, Italy): 20 years of monitoring (1995–2014). Rend Fis Acc Lincei 28:317–328. https://doi.org/10.1007/s12210-016-0557-z

Doneen LD (1964) Water quality for agriculture. Department of Agriculture, University of California, Davis, p 48

Durov SA (1948) Natural waters and graphic representation of their composition. Dokl Akad Nauk SSSR 59:87–90

Eaton EM (1950) Significance of carbonate in irrigation water. Soil Sci 69:12–133. https://doi.org/10.1097/00010694-195002000-00004

Ersoy AF, Karaca Z (2019) Determination of groundwater parameters for drinking and agricultural use in the coastal region of Engiz Aquifer System, Samsun (Turkey). Arab J Geosci 12:198. https://doi.org/10.1007/s12517-019-4365-5

Gaillardet J, Dupre B, Louvat P, Allegre CJ (1999) Global silicate weathering and CO_2 consumption rate deduced from the chemistry of large rivers. Chem Geology 59:3–30. https://doi.org/10.1016/S0009-2541(99)00031-5

Gibbs RJ (1970) Mechanism controlling world water chemistry. Science 17:1088–1090

Goyal D, Haritash AK, Singh SK (2021) A comprehensive review of groundwater vulnerability assessment using index-based, modelling and coupling methods. J Environ Manage 296:113161. https://doi.org/10.1016/j.jenvman.2021.113161

Haritash AK, Kaushik CP, Kaushik A, Kansal A, Yadav AK (2008) Suitability assessment of groundwater in some villages of Rewari district in Haryana. Environ Monit Assess 145(1–3):397–406. https://doi.org/10.1007/s10661-007-0048-x

Haritash AK, Shan V, Singh P, Singh SK (2015) Preliminary investigation of environmental status of Bhindawas Bird Sanctuary. Intern J Eng Res Technol 4(03):53–56

Haritash AK, Mathur K, Singh P, Singh SK (2017) Hydrochemical characterization and suitability assessment of groundwater in Baga-Calangute stretch of Goa. India Environ Earth Sci 76:341. https://doi.org/10.1007/s12665-017-6679-5

Haritash AK, Aggarwal A, Soni J, Sharma K, Sapra M, Bhupinder S (2018) Assessment of fluoride in groundwater and urine, and prevalence of fluorosis among school children in Haryana. India Appl Water Sci 8:52. https://doi.org/10.1007/s13201-018-0691-0

Karaca O, Claudio C, Reddy KR (2017) Acid Pond sediment and mine tailings contaminated with metals: physicochemical characterization and electrokinetic remediation. Environmental Earth Science 76(12):408. https://doi.org/10.1007/s12665-017-6736-0

Kaushik A, Kansal A, Meena S, Kumari S, Kaushik CP (2009) Heavy metal contamination of river Yamuna, Haryana, India: assessment by metal enrichment factor of the sediments. J Hazard Mater 164:265–270. https://doi.org/10.1016/j.jhazmat.2008.08.031

Kazmi AA, Bhatia A, Shaida A, Sharma M, Starkl M, Trivedi RC (2013) A short screening study on water quality of Indian rivers and lakes. J Indian Water Resourc Soc 33(3):28–33

Kelly WP (1963) Use of saline irrigation water. Soil Sci 95:355–391

Kumar S, Lata S, Yadav J, Yadav JP (2017) Relationship between water, urine and serum fluoride and fluorosis in school-children of Jhajjar District, Haryana, India. Appl Water Sci 7:3377–3384. https://doi.org/10.1007/s13201-016-0492-2

Langelier WF (1936) The Analytical Control of Anti-Corrosive Water Treatment. J Am Water Works Assoc 28(10):1500

Meybeck M (1986) Compositiondes ruisseau non pollutes de France. Sci Geo Bull 39:3–77

Paliwal KV 1972 Irrigation with saline water. IARI Monograph No. 2 (New Series), New Delhi; 198

Pham TL, Tran THY, Tran TT, Ngo XQ, Nguyen XD (2022) Assessment of surface water quality in a drinking water supply reservoir in Vietnam: a combination of different indicators. Fis. Acc. Lincei, Rend. https://doi.org/10.1007/s12210-022-01086-5

Pipil H, Haritash AK, Reddy KR (2021) Seasonal variability and kinetics of phosphate removal in a Phragmites-based engineered wetland. Rend Fis Acc Lincei 32:729–735. https://doi.org/10.1007/s12210-021-01017-w

Raman V (1985) Impact of corrosion in the conveyance and distribution of water. J Indian Water Works Assoc 15(11):115–121

Rawal I, Joshi H, Chaudhary BH (2014) Study of water quality of lake Pichhola of Udaipur (India). J Herb Med Toxicol 8(1):72–75

Rossi D, Romano E, Guyennon N, Rainaldi M, Ghergo G, Mecali A, Parrone D, Taviani S, Scala A, Perugini E (2019) The present state of Lake Bracciano: hope and despair. Rend Fis Acc Lincei 30:83–91. https://doi.org/10.1007/s12210-018-0733-4

Ryner JW (1944) a new index for determining amount of calcium carbonate scale formed by water. J Amer Wat Assn 36:472–486

Ryznar JW (1944) An index for determining amount of calcium carbonate scale formed by water. J Am Water Works Assoc 36(4):472–486

Shan V, Singh SK, Haritash AK (2020) Evaluation of water quality and potential metal contamination in ecologically important Bhindawas bird sanctuary. India Applied Water Science 11(8):1–9. https://doi.org/10.1007/s13201-020-01334-9

Singh P, Haritash AK (2019) An Active biological food chain can restrict Eutrophication: case study of a pond ecosystem. Indian J Waste Manag 3(1):15–20

Subramanian V (1987) Environmental geochemistry of Indian River basins-a review. J Geol Soc India 29:205–220

Vashistha J, John PJ, Paulose PV (2016) Water Quality Analysis of Siliserh, India. Intern J Curr Microbiol Appl Sci. 5(10):439–450

WHO (2006) Guidelines for drinking water quality, First addendum to, 3rd edn. World Health Organization, Geneva

Wilcox LV (1955) Classification and use of irrigation water. USDA Circular 969, Washington DC. 19.

Yadav A, Kataria A, Singh K, Mathur K, Goswami S, Haritash AK (2015) Seasonal assessment of trophic state of a palustrine water body. Intern J Eng Res Technol 4(03):37–40

Yun YJ, An KG (2016) Roles of N: P ratios on trophic structures and ecological stream health in lotic ecosystems. Water 8(1):22

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

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

Authors and Affiliations

Chitrakshi[1] · Anil Kumar Haritash[1]

[1] Department of Environmental Engineering, Delhi Technological University, Bawana Road, Shahbad Daulatpur, Delhi 110042, India

Mathematical Foundations of Computing

doi:10.3934/mfc.2022034

APPROXIMATION BY DURRMEYER VARIANT OF CHENEY-SHARMA CHLODOVSKY OPERATORS

Chandra Prakash[*1], Durvesh Kumar Verma[2] and Naokant Deo[1]

[1]Department of Mathematics, Delhi Technological University, Delhi-110042, India

[2]Department of Mathematics, Miranda House, University of Delhi, Delhi-110007, India

(Communicated by Tuncer Acar)

Abstract. In this paper, we are dealing with Cheney-Sharma Chlodovsky Durrmeyer operators and studying their approximation properties. The Bohman-Korovkin theorem is verified and estimated the convergence properties using of modulus of continuity, Lipschitz- type space, and Ditzian-Totik modulus of continuity. After that, the weighted approximation result is also given. Finally, some results related to the A-statistical convergence of the operators are obtained.

1. **Introduction.** The theory of approximation is connected with how real functions and polynomials can best be approximated with simpler are trigonometric or algebraic functions. The Weierstrass approximation theorem plays a significant role in various branches of mathematics, sciences and engineering. It states that every continuous function on a closed interval can be uniformly approximated by a sequence of polynomials. In 1912, S. N. Bernstein proposed an operator that plays a vital role in approximation theory which is known as Bernstein polynomial. With the help of these operators, he gave constructive proof of the Weierstrass approximation theorem on the closed interval $[0,1]$. Later, many authors gave several local and global approximation results, for details we refer the readers to ([30, 16, 10, 7, 8]).

In 1967, Durrmeyer [19] gave an integral modification of the Bernstein polynomials with the help of Lebesgue integrable function on the interval [0,1] as:

$$D_n(h,x) = (n+1)\sum_{\nu=0}^{n} \vartheta_{n,\nu}(x) \int_0^1 \vartheta_{n,\nu}(t)h(t)dt, \tag{1}$$

where $0 \leq x \leq 1,\ h \in [0,1]$ and $\vartheta_{n,\nu}(x) = \binom{n}{\nu} x^{\nu}(1-x)^{n-\nu}$.

The operators (1) are known as the Bernstein-Durrmeyer operators. The Bernstein polynomials and its generalized operators have been extensively studied by several researchers. These operators have many important applications in the field of Mathematics, computer science, and graphic design. For the more information about the Durrmeyer operators and their properties, readers should refer the following articles (cf.[5, 29, 23, 25, 36, 31, 9, 32, 28]).

2020 *Mathematics Subject Classification.* Primary: 26A15, 41A25; Secondary: 40A35.

Key words and phrases. Cheney-Sharma-Chlodovsky operators, Lipschitz-type space, weighted approximation, modulus of continuity, A-statistical approximation.

*Corresponding author: Chandra Prakash.

In 1932, Chlodovsky [15] introduced the generalization of Bernstein polynomial known as classical Bernstein-Chlodovsky polynomials on an unbounded interval. These polynomials $C_n : C[0,\infty) \to C[0,\infty)$, $n \in \mathbb{N}$ is defined as

$$C_n(h,x) = \begin{cases} \sum_{\nu=0}^{n} h\left(\frac{\nu}{n}b_n\right)\binom{n}{\nu}\left(\frac{x}{b_n}\right)^{\nu}\left(1-\frac{x}{b_n}\right)^{n-\nu}, 0 \le x \le b_n \\ h(x), \qquad\qquad x > b_n, \end{cases}$$

where $h \in C[0,\infty)$, $x \in [0,\infty)$ and $0 \le x \le b_n$ and $\{b_n\}$ is a positive sequence with $\lim_{n\to\infty} b_n = \infty$, and $\lim_{n\to\infty} \frac{b_n}{n} = 0$.

In [1], the eminent Norwegian mathematician Niels Henrik Abel gave famous equality as:

$$(x+y)^n = \sum_{\nu=0}^{n}\binom{n}{\nu}x(x-\nu\beta)^{\nu-1}(y+\nu\beta)^{n-\nu},\ \beta \ge 0.$$

Use of above equality we mention more Abel type formulas

$$(x+y+n\beta)^n = \sum_{\nu=0}^{n}\binom{n}{\nu}x(x+\nu\beta)^{\nu-1}(y+(n-\nu)\beta)^{n-\nu},$$

$$(x+y+n\beta)^n = \sum_{\nu=0}^{n}\binom{n}{\nu}(x+\nu\beta)^{\nu}y(y+(n-\nu)\beta)^{n-\nu-1},$$

$$(x+y+n\beta)^{n-1} = \sum_{\nu=0}^{n}\binom{n}{\nu}x(x+\nu\beta)^{\nu-1}y[y+(n-\nu)\beta]^{n-\nu-1},$$

$$(x+y)(x+y+n\beta)^{n-1} = \sum_{\nu=0}^{n}\binom{n}{\nu}x(x+\nu\beta)^{\nu-1}y[y+(n-\nu)\beta]^{n-\nu-1},$$

where $x, y \in \mathbb{R}$.

Using these equalities, Cheney-Sharma [14] in 1964 introduced a generalization of Bernstein polynomial as follows:

$$\Psi_n(h,x) = (1+n\beta)^{1-n}\sum_{\nu=0}^{n} h\left(\tfrac{\nu}{n}\right)\binom{n}{\nu}x(x+k\beta)^{\nu-1}(1-x)[1-x+(n-\nu)\beta]^{n-\nu-1}. \quad (2)$$

For $\beta = 0$ above operators reduce to the classical Bernstein operators.

Recently, Söylemez and Taşdelen [33] proposed Cheney-Sharma-Chlodovsky operators depending on some parameters as follows:

$$G_n(h,x) = (1+n\beta)^{1-n}\sum_{\nu=0}^{n} h\left(\frac{\nu}{n}b_n\right)\binom{n}{\nu}\frac{x}{b_n}\left(\frac{x}{b_n}+\beta\right)^{\nu-1}\left(1-\frac{x}{b_n}\right) \times\left[1-\frac{x}{b_n}+(n-\nu)\beta\right]^{n-\nu-1} \quad (3)$$

for $0 \le x \le b_n$ and $f(x)$ for $x > b_n$. Here sequence $\{b_n\}$ is the same as defined in classical Bernstein-Chlodovsky polynomials. They proved the order of convergence and discuss some important results of these operators. For $b_n = 1$ the operators (3) reduce to (2).

In this paper, encouraged by the above work, we introduced a new type of Cheney-Sharma-Chlodovsky -Durrmeyer operators, with the help of Lebesgue integrable function on $[0, b_n]$ as

$$\tilde{D}_n(h,x) = \frac{(n+1)}{b_n}\sum_{\nu=0}^{n} \wp_{n,\nu,\beta}\left(\frac{x}{b_n}\right)\int_0^{b_n} \vartheta_{n,\nu}\left(\frac{x}{b_n}\right)h(x)dx, \tag{4}$$

where $0 \leq x \leq b_n$, $\lim_{n\to\infty} b_n \to \infty$, $\lim_{n\to\infty} \frac{b_n}{n} = 0$,
$\wp_{n,\nu,\beta}\left(\frac{x}{b_n}\right) = (1+n\beta)^{1-n}\binom{n}{\nu}\frac{x}{b_n}\left(\frac{x}{b_n}+\nu\beta\right)^{\nu-1}\left(1-\frac{x}{b_n}\right)\left[1-\frac{x}{b_n}+(n-\nu)\beta\right]^{n-\nu-1}$,
and

$$\vartheta_{n,\nu}\left(\frac{x}{b_n}\right) = \binom{n}{\nu}\left(\frac{t}{b_n}\right)^{\nu}\left(1-\frac{t}{b_n}\right)^{n-\nu}.$$

The aim of this paper is to study some approximation properties of the operators (4). In section 2, we have given some lemmas and auxiliary results to prove our main results. In the next section, we verified the Bohman-Korovkin theorem and obtained convergence properties in terms of modulus of continuity, Lipschitz-type space and Ditzian-Totic modulus of continuity. In section 4, we have estimated the weighted approximation result. In the last section, the A-statistical convergence properties of the operators (4) are discussed.

2. **Auxiliary results.** Moments and Central moments of the Cheney-Sharma Chlodovsky and Cheney-Sharma Chlodovsky Durrmeyer operators are given by

Lemma 2.1. [33] *The moments of the operators $G_n((e_r(t)=t^r),x)$, $r=\overline{0,2}$, we have*

$$\begin{aligned}
G_n(e_0,x) &= 1,\\
G_n(e_1,x) &= x,\\
G_n(e_2,x) &\leq x^2(1+2n\beta)\\
&\quad +x\left(\frac{b_n}{n}+2n\beta^2 b_n+n^2\beta^3 b_n+2\beta(1+n\beta)b_n+n\beta^2(1+n\beta)b_n\right).
\end{aligned}$$

Lemma 2.2. *The moments of the operators $\tilde{D}_n(t^r,x)$, $r=\overline{0,2}$, we have*

$$\begin{aligned}
\tilde{D}_n(e_0,x) &= 1,\\
\tilde{D}_n(e_1,x) &= \frac{nx+b_n}{n+2},\\
\tilde{D}_n(e_2,x) &\leq \frac{n^2x^2(1+2n\beta)+nx(4+2n\beta+5n^2\beta^2+2n^3\beta^3)b_2+2b_n^2}{(2+n)(3+n)}.
\end{aligned}$$

Proof. After simple computations, we find a relation of the given operators (4)

$$\tilde{D}_n(t^r,x) = \frac{(n+1)!}{(n+r+1)!}b_n^r\sum_{\nu=0}^{n}\wp_{n,\nu,\beta}\left(\frac{x}{b_n}\right)\frac{(\nu+r)!}{\nu!}. \tag{5}$$

Substitute $r=0$ in the equation (5), we obtain the result $\tilde{D}_n(e_0,x)=1$.

Again, take $r=1$ in the equation (5), we get

$$\tilde{D}_n(e_1,x) = \frac{(n+1)!}{(n+2)!}b_n\sum_{\nu=0}^{n}\wp_n\left(\frac{x}{b_n}\right)\frac{(\nu+1)!}{\nu!}$$

$$\begin{aligned}
&= \frac{b_n}{(n+2)}\left\{\sum_{\nu=0}^{n}\left(\frac{\nu}{n}b_n\right)\frac{n}{b_n}\wp_n\left(\frac{x}{b_n}\right)+\sum_{\nu=0}^{n}\wp_n\left(\frac{x}{b_n}\right)\right\}\\
&= \frac{n}{(n+2)}\sum_{\nu=0}^{n}\left(\frac{\nu}{n}b_n\right)\wp_n\left(\frac{x}{b_n}\right)+\frac{b_n}{(n+2)}\sum_{\nu=0}^{n}\wp_n\left(\frac{x}{b_n}\right)\\
&= \frac{n}{(n+2)}G_n(t,x)+\frac{b_n}{(n+2)}G_n(1,x)\\
&= \frac{n}{(n+2)}x+\frac{b_n}{(n+2)},
\end{aligned}$$

thus

$$\tilde{D}_n(e_1,x)=\frac{nx+b_n}{n+2}.$$

Now, take $r=2$ in the equation (5), we find the result

$$\begin{aligned}
\tilde{D}_n(e_2,x) &= \frac{(n+1)!}{(n+3)!}b_n^2\sum_{\nu=0}^{n}\wp_n\left(\frac{x}{b_n}\right)\frac{(\nu+2)!}{\nu!}\\
&= \frac{b_n^2}{(n+2)(n+3)}\left\{\sum_{\nu=0}^{n}\nu^2\wp_n\left(\frac{x}{b_n}\right)\right.\\
&\quad\left.+\sum_{\nu=0}^{n}3\nu\wp_n\left(\frac{x}{b_n}\right)+\sum_{\nu=0}^{n}2\wp_n\left(\frac{x}{b_n}\right)\right\}\\
&= \frac{b_n^2}{(n+2)(n+3)}\left\{\sum_{\nu=0}^{n}\left(\frac{\nu^2}{n^2}b_n^2\right)\left(\frac{n^2}{b_n^2}\right)\wp_n\left(\frac{x}{b_n}\right)\right.\\
&\quad+3\sum_{\nu=0}^{n}\left(\frac{\nu}{n}b_n\right)\left(\frac{n}{b_n}\right)\wp_n\left(\frac{x}{b_n}\right)\\
&\quad\left.+2\sum_{\nu=0}^{n}\wp_n\left(\frac{x}{b_n}\right)\right\}\frac{n^2}{(n+2)(n+3)}G_n(t^2,x)\\
&= +\frac{3nb_n}{(n+2)(n+3)}G_n(t,x)+\frac{2b_n^2}{(n+2)(n+3)}G_n(1,x)\\
&\leq \frac{n^2x^2(1+2n\beta)+nx(4+2n\beta+5n^2\beta^2+2n^3\beta^3)b_2+2b_n^2}{(2+n)(3+n)}.
\end{aligned}$$

Hence,

$$\tilde{D}_n(e_2,x)\leq\frac{n^2x^2(1+2n\beta)+nx(4+2n\beta+5n^2\beta^2+2n^3\beta^3)b_2+2b_n^2}{(2+n)(3+n)}.$$

□

The Central moment of the operators $\tilde{D}_n((t-x)^2,x)$, we have

$$\begin{aligned}
&\tilde{D}_n((t-x)^2,x)\\
&\leq\frac{x^2(6-n+2n^3\beta)+x(-6+2n+2n^2\beta+5n^3\beta^2+2n^4\beta^3)n_n+2b_n^2}{(2+n)(3+n)}.
\end{aligned}$$

Lemma 2.3. *Let f be a real-valued continuous bounded function on $[0,1]$, then*

$$|\tilde{D}_n(f)|\leq\|f\|.$$

Where sup-norm $\|.\|$ *is defined as* $\|f\| = \sup_{x\in[0,b_n]} |f(x)|$.

3. **Local approximation.** In the present section, we verified the Bohman-Korovkin theorem and estimated local approximation in terms of modulus of continuity, Lipschitz-type space, and Ditzian-totik modulus of smoothness.

Theorem 3.1. *Let* $\tilde{D}_n$ *be the operators given by (5) and* $h \in C[0,\infty) \cap E$, *then the following relations holds*

$$\lim_{n\to\infty} \tilde{D}_n(h,x) = h(x)$$

uniformly on each compact subset of $[0,\infty)$, *where*

$$E = \left\{ h : x \in [0,\infty), \frac{h(x)}{1+x^2} \text{ is convergent as } x \to \infty \right\}.$$

Here β *as a sequence of positive real numbers such that* $\beta = \beta_n$ *and* $\lim_{n\to\infty} n\beta_n = 0$.

Proof. It is sufficient to prove that the operators $\tilde{D}_n$ verify the conditions, use of Korovkin's theorem

$$\lim_{n\to\infty} \tilde{D}_n(e_r,x) = x^r,\ r = \overline{0,2}$$

uniformly on each compact subset of $[0,\infty)$. From the Lemma 2.2,

$$\tilde{D}_n(e_0,x) = 1 \text{ for every } n \in \mathbb{N}$$

tends to 1,

$$\tilde{D}_n(e_1,x) = \frac{nx+b_n}{n+2}$$

tends to x, and

$$\tilde{D}_n(e_2,x) \leq \frac{n^2x^2(1+2n\beta) + nx(4+2n\beta+5n^2\beta^2+2n^3\beta^3)b_2 + 2b_n^2}{(2+n)(3+n)}$$

tends to x^2 as n large. □

In what follows we will show the appraisal of the Cheney-Sharma Chlodovsky Durrmeyer operators on the specific Lipschitz-type space. In [27], the following is defined as

$$Lip^*_{\aleph}(\sigma) = \left\{ h \in C[0,1] : |h(t)-h(x)| \leq \aleph \frac{|t-x|^\sigma}{(t+x)^{\frac{\sigma}{2}}} \right\},$$

where $x \in (0,1), t \in [0,1]$, and $0 < \sigma \leq 1$.

In the following theorem, we obtain the rate of convergence of the operators $\tilde{D}_n$ for functions in $Lip^*_{\aleph}(\sigma)$.

Theorem 3.2. *Let* $h \in Lip^*_{\aleph}(\sigma)$ *and* $0 < \sigma \leq 1$. *Then for each* $x \in (0,\infty)$, *we have*

$$|\tilde{D}_n(h(t);x) - h(x)| \leq \aleph \left(\frac{\tilde{D}_n((t-x)^2;x)}{x} \right)^{\frac{\sigma}{2}}.$$

Proof. Let $h \in Lip^*_{\aleph}(\sigma)$ and $x \in (0,\infty)$, $t \in (0,\infty)$, we have

$$\begin{aligned} |\tilde{D}_n(h(t);x) - h(x)| &\leq \tilde{D}_n(|h(t)-h(x)|;x) \\ &\leq \aleph.\tilde{D}_n\left(\frac{|t-x|^\sigma}{(t+x)^{\frac{\sigma}{2}}};x \right) \end{aligned}$$

$$\leq \frac{\aleph}{x^{\frac{\sigma}{2}}}\tilde{D}_n(|t-x|^{\sigma};x). \tag{6}$$

Taking $p=\frac{2}{\sigma}$ and $q=\frac{2}{2-\sigma}$ and applying Hölder's inequality, we obtain

$$\begin{aligned}\tilde{D}_n(|t-x|^{\sigma};x) &\leq \left\{\tilde{D}_n(|t-x|^2;x)\right\}^{\frac{\sigma}{2}}.\left\{\tilde{D}_n(1^{\frac{2}{2-\sigma}};x)\right\}^{\frac{2-\sigma}{2}}\\ &\leq \left\{\tilde{D}_n(|t-x|^2;x)\right\}^{\frac{\sigma}{2}}. \end{aligned}\tag{7}$$

Using (6) and (7), we get the required result

$$|\tilde{D}_n(h(t);x)-h(x)| \leq \aleph\left(\frac{\tilde{D}_n((t-x)^2;x)}{x}\right)^{\frac{\sigma}{2}}.$$

□

The Peetre K-functional is defined by

$$K_2(h,\delta)=\inf\{\|h-g\|+\delta\|g''\| : g\in W^2\},\ \delta>0$$

where

$$W^2=\{g\in C[0,1] : g', g''\in C[0,1]\}$$

and $\|.\|$ is the uniform on $C[0,1]$.

$$K_2(h,\delta)\leq C\omega_2(h,\sqrt{\delta}), \tag{8}$$

where the second order modulus of smoothness for $h\in C[0,1]$ is defined as

$$\omega_2(h,\sqrt{\delta})=\sup_{0<l\leq\sqrt{\delta},\{x,x+2l\}\in[0,1]}|h(x+2l)-2h(x+l)+h(x)|.$$

The usual modulus of continuity for $h\in C[0,1]$ is defined as

$$\omega(h,\delta)=\sup_{0<l\leq\sqrt{\delta},\{x,x+l\}\in[0,1]}|h(x+l)-h(x)|.$$

Theorem 3.3. *Let $h\in C_B[0,\infty)$, we have*

$$|\tilde{D}_n(h,x)-h(x)|\leq C\omega_2(h,\sqrt{\delta})+\omega\left(h,\frac{b_n}{n+2}\right),$$

where C is a positive constant and $\delta=|\tilde{D}_n((t-x)^2,x)|+\left(\frac{b_n}{n+2}\right)$.

Proof. We define operators $H_n(t,x)$ as follows

$$H_n(h,x)=\tilde{D}_n(t,x)-h\left(\frac{nx+b_n}{n+2}\right)+h(x).$$

These operators are linear and preserve the linear functions in view of Lemma 2.1. Let $g\in W^2$ from the Taylors expansion of g, we have

$$g(t)=g(x)+(t-x)g'(x)+\int_x^t(t-u)g''(u)du.$$

Applying the operator H_n on above, we obtain

$$H_n(g,x)=g(x)+g'(x)H_n((t-x),x)+H_n\left(\int_x^t(t-u)g''(u)du,x\right)$$

or

$$\begin{aligned}
|H_n(g,x)-g(x)| &= |H_n\left(\int_x^t (t-u)g''(u)du, x\right)| \\
&\leq |\tilde{D}_n\left(\int_x^t (t-u)g''(u)du, x\right)| \\
&\quad + \left|\int_x^{\frac{nx+b_n}{n+2}} \left(\frac{nx+b_n}{n+2}-u\right) g''(u)du\right| \\
&\leq \left\{\left|\tilde{D}_n\left(\int_x^t (t-u)du, x\right)\right| \right. \\
&\quad \left. + \left|\int_x^{\frac{nx+b_n}{n+2}} \left(\frac{nx+b_n}{n+2}-u\right) du\right|\right\} \|g''\| \\
&\leq \left\{|\tilde{D}_n((t-x)^2,x)| + \left(\frac{b_n}{n+2}\right)^2\right\} \|g''\| \\
&= \delta\|g''\|,
\end{aligned}$$

where $\delta = |\tilde{D}_n((t-x)^2,x)| + \left(\frac{b_n}{n+2}\right)^2$.

Now

$$\begin{aligned}
|\tilde{D}_n(h,x)-h(x)| &= |H_n(h-g,x)-(h-g)(x)| + |H_n(g,x)-g(x)| \\
&\quad + \left|h\left(\frac{nx+b_n}{n+2}\right)-h(x)\right| \\
&\leq 2\|h-g\| + \delta\|g''\| + \omega\left(h, \frac{b_n}{n+2}\right).
\end{aligned}$$

Taking infimum over all $g \in W^2$, we get

$$|\tilde{D}_n(h,x)-h(x)| \leq K_2(h,\delta) + \omega\left(h, \frac{b_n}{n+2}\right).$$

In view of (8), we obtain the result

$$|\tilde{D}_n(h,x)-h(x)| \leq C\omega_2(h,\sqrt{\delta}) + \omega\left(h, \frac{b_n}{n+2}\right).$$

□

4. **Weighted Korovkin's theorem.** In this section, we obtained the weighted result of the given operators.

Consider the class of function $C_\rho^k(\mathbb{R}^+) = \left\{h \in C_\rho(\mathbb{R}^+);\ \lim_{x\to\infty} \frac{h(x)}{\rho(x)} = k_h\right\}$ and the norm $\|h\|_\rho = \sup_{x\in\mathbb{R}^+} \frac{|h(x)|}{\rho(x)}$. Where k_h is a constant depending on h and $\rho(x) = 1+x^2$.

Theorem 4.1. *Suppose that $\{b_n\}$ and $\{\beta_n\}$ are sequence of positive numbers such that $\lim_{n\to\infty} b_n = \infty$, $\lim_{n\to\infty} \frac{b_n}{n} = 0$ and $\lim_{n\to\infty} n\beta_n = 0$, for each $h \in C_\rho^k(\mathbb{R}^+)$ and $n \in \mathbb{N}$, then*

$$\lim_{n\to\infty} \|\tilde{D}_n(h) - h\|_\rho = 0.$$

Proof. Using Lemma 2.2, it is sufficient to verify the following conditions

$$\|\tilde{D}_n(t^r,x)-x^r\|_\rho=0,\ \ r=\overline{0,2}. \tag{9}$$

Since $\lim\limits_{n\to\infty}\|\tilde{D}_n(1,x)-1||_\rho=0$, for $r=0$, (9) holds. And $\lim\limits_{n\to\infty}\|\tilde{D}_n(t,x)-x||_\rho=0$, for $r=1$, also holds.

Again,

$$\begin{aligned}\|\tilde{D}_n(t,x)-x^2||_\rho \quad\leq\quad & \sup\left\{\left|\frac{n^2x^2(1+2n\beta)}{(1+x^2)(2+n)(3+n)}\right|\right.\\ &+\left|\frac{nx(4+2n\beta+5n^2\beta^2+2n^3\beta^3)b_n}{(1+x^2)(2+n)(3+n)}\right|\\ &\left.+\left|\frac{2b_n^2}{(1+x^2)(2+n)(3+n)}\right|\right\}\end{aligned}$$

Hence

$$\lim_{n\to\infty}\|\tilde{D}_n(t^2,x)-x^2||_\rho=0.$$

the condition (9) holds for $r=2$. □

5. **A-statistical approximation.** In this section, we estimate the $A-$statistical convergence of the given operators $\tilde{D}_n$ to identity operators on the weighted spaces.

Let $A=(a_{n\nu})$ be a summability matrix and let $x=(x_\nu)$ be a sequence. If the series

$$(Ax)_n=\sum_\nu a_{n\nu}x_\nu$$

is convergent for each n then, $Ax=\{(Ax)_n\}$ is the $A-$transformation of x. If the sequence Ax converges to a constant M then, we say that x is a $A-$summable to M. A summability matrix A is said to be regular if $\lim_n(Ax)_n=M$, whenever $\lim\limits_\nu x_\nu=M$ [11].

Consider $A=(a_{n\nu})$ be a nonnegative regular summability matrix and let J be a subset of positive integer. Then J is said to be $A-$density $\delta_A(J)$ if the limit exist [12, 13], where

$$\delta_A(J)=\lim_n\sum_{\nu\in J}a_{n\nu}.$$

The sequence $x=(x_\nu)$ is said to be $A-$statistically convergent to real number γ if for any $\epsilon>0$ and

$$\lim_\nu\sum_{\nu:|x_\nu-\gamma|\geq\epsilon}a_{n\nu}=0.$$

In this case, we write $st_A-\lim x=\gamma$ [20, 22]. We have x is $A-$statistically convergent to γ if and only if for any $\epsilon>0,\delta_A(J_\epsilon)=0$, where $J_\epsilon=\{\nu\in\mathbb{N}:|x_\nu-\gamma|\geq\epsilon\}$. If A is the identity matrix I, then $I-$statistical convergence reduces to ordinary convergence and if $A=C_1$, the Cesáro matrix of order one, then it coincides with statistical convergence.

Firstly we discuss the weighted Korovkin type approximation theorem for the $A-$statistical convergence given by Duman and Orhan [18] in 2004 and, Dilek and Mehmet [34] in 2017.

Theorem 5.1. *Let $A = (a_{n\nu})$ be a nonnegative regular summability matrix and $\overline{\varrho}_1, \overline{\varrho}_2$ weight functions such that*

$$\lim_{|x|\to\infty} \frac{\overline{\varrho}_1(x)}{\overline{\varrho}_2(x)} = 0. \tag{10}$$

Assume that $(T_n)_{n\geq 1}$ is a sequence of positive linear operators from $C_{\overline{\varrho}_1}(\mathbb{R})$ into $B_{\overline{\varrho}_2}(\mathbb{R})$. Now $st_A - \lim_n \|T_n h - h\|_{\overline{\varrho}_1}$, for all $f \in C_{\overline{\varrho}_1}(\mathbb{R})$ iff

$$st_A - \lim_n \|T_n h_\upsilon - h_\upsilon\|_{\overline{\varrho}_1} = 0, \ \upsilon = 0, 1, 2,$$

where

$$h_\upsilon(x) = \frac{x^\upsilon \overline{\varrho}_1(x)}{1+x^2}, \quad \upsilon = 0, 1, 2.$$

Corollary 5.2. [18]. *Let $A = (a_{n\nu})$ be a nonnegative regular summability matrix and let (L_n) be a sequence of positive linear operators acting from $C_{\rho_o}(R^+)$ into $B_{\rho_\lambda}(R^+)$, $\lambda > 0$ one has*

$$st_A - \lim_n \|L_n h - h\|_{\rho_\lambda} = 0, \quad h \in C_{\rho_o}(R^+),$$

iff

$$st_A - \lim_n \|L_n e_i - e_i\|_{\rho_o} = 0, \quad i = 0, 1, 2, \tag{11}$$

where $\rho_o(x) = 1 + x^2$ and $\rho_\lambda(x) = 1 + x^{2+\lambda}$, $\lambda > 0$.

Using the above results we will estimate the korovkin type statical theorem for given operators.

Theorem 5.3. *Let (b_n) and (β_n) be a sequences of positive numbers such that $st_A - \lim_{n\to\infty} \frac{b_n}{n} = 0$, $st_A - \lim_{n\to\infty} n\beta_n = 0$ and $A = (a_{nk})$ be a nonnegative regular summability matrix. Then for each $h \in C^k_\rho(\mathbb{R}^+)$, we have*

$$st_A - \lim_{n\to\infty} \|\tilde{D}_n h - h\|_{\rho_\lambda} = 0,$$

where $\rho_o(x) = 1 + x^2$ and $\rho_\lambda(x) = 1 + x^{2+\lambda}$, $\lambda > 0$.

Proof. From the above results, it is sufficient to prove that the operators $\tilde{D}_n$ verify (9), we get

$$st_A - \lim_{n\to\infty} \|\tilde{D}_n(e_0; .) - e_0\|_{\rho_o} = 0$$

and

$$st_A - \lim_{n\to\infty} \|\tilde{D}_n(e_1; .) - e_1\|_{\rho_o} = 0.$$

Now,

$$\begin{aligned} \|\tilde{D}_n(e_2) - x^2\|_{\rho_o} &\leq \sup\left\{ \left| \frac{x^2}{1+x^2} \frac{(2\beta n^3 - 5n - 6)}{(2+n)(3+n)} \right| \right. \\ &\quad + \left| \frac{x}{1+x^2} \frac{n(4 + 2n\beta + 5n^2\beta^2 + 2n^3\beta^3)b_2}{(2+n)(3+n)} \right| \\ &\quad \left. + \left| \frac{1}{1+x^2} \frac{2b_n^2}{(2+n)(3+n)} \right| \right\} \\ &= \frac{2\beta n^3 - 5n - 6}{(2+n)(3+n)} + \frac{n(4 + 2n\beta + 5n^2\beta^2 + 2n^3\beta^3)b_2}{(2+n)(3+n)} \end{aligned}$$

$$\begin{aligned} &+\frac{2b_n^2}{(2+n)(3+n)} \\ = \;& J_n. \end{aligned}$$

For a given $\epsilon > 0$, we define the following sets

$$\begin{aligned} N &= \{\nu : \|\tilde{D}_n(e_2;.) - e_2\|_{\rho_o} \geq \epsilon\}, \\ N_1 &= \left\{\nu : \frac{2\beta n^3 - 5n - 6}{(2+n)(3+n)} \geq \frac{\epsilon}{3}\right\}, \\ N_2 &= \left\{\nu : \frac{n(4+2n\beta+5n^2\beta^2+2n^3\beta^3)b_2}{(2+n)(3+n)} \geq \frac{\epsilon}{3}\right\}, \\ N_3 &= \left\{\nu : \frac{2b_n^2}{(2+n)(3+n)} \geq \frac{\epsilon}{3}\right\}. \end{aligned}$$

Then, we see that $N \subseteq N_1 \cup N_2 \cup N_3$. Therefore, we get

$$\sum_{n:\|\tilde{D}_n(e_2;.)-e_2\|_{\rho_o}\geq 0} a_{n\nu} \leq \sum_{n\in N_1} a_{n\nu} + \sum_{n\in N_2} a_{n\nu} + \sum_{n\in N_3} a_{n\nu},$$

taking the limit $\nu \to \infty$ in above, we get the result
$st_A - \lim_n \|\tilde{D}_n(e_2;.) - e_2\|_{\rho_o} = 0$. □

Acknowledgments. The authors are thankful to the referees for valuable suggestions, leading to an overall improvement in the paper.

REFERENCES

[1] N. H. Abel, *Démonstration dúne expression de laquelle la formule biñome est un cas particulier*, *Journ. f̈ur Reine und Angewendte Mathematik*, **1** (1826), 159-160, Oeuvres Complètes, Christiania, Groendahls, 1839.

[2] U. Abel, V. Gupta and M. Ivan, Asymptotic approximation of functions and their derivatives by generalized Baskakov-Szász-Durrmeyer operators, *Anal. Theory Appl.*, **21** (2005), 15-26.

[3] U. Abel, V. Gupta and R. N. Mohapatra, Local approximation by a varient of Bernstein-Durrmeyer operators, *Nonlinear Anal. Theory Methods Appl. Ser. A: Theory Methods*, **68** (2008), 3372-3381.

[4] U. Abel and M. Heilmann, The complete asympotic expansion for Bernstein-Durrmeyer operators with Jacobi weights, *Mediterr. J. Math.*, **1** (2004), 487-499.

[5] U. Abel, M. Ivan and Z. Li, Local approximation by generalized Baskakov-Durrmeyer operators, *Numer. Funct. Anal. Optimiz.*, **28** (2007), 245-264.

[6] U. Abel, M. Ivan and R. Păltănea, The Durrmeyer variant of an operator defined by D. D. Stancu, *Appl. Math. Comput.*, **259** (2015), 116-123.

[7] T. Acar, M. C. Montano, P. Garrancho and V. Leonessa, On Bernstein-Chlodovsky operators preserving e^{-2x}, *Bull. Belg. Math. Soc. Simon Stevin*, **26** (2019), 681-698.

[8] T. Acar, M. C. Montano, P. Garrancho and V. Leonessa, Voronovskaya type results for Bernstein-Chlodovsky operators preserving e^{-2x}, *J. Math. Anal. Appl.*, **491** (2020), 124307.

[9] P. N. Agrawal and J. K. Singh, Better approximation by a Durrmeyer variant of $\alpha-$ Baskakov operators, *Mathematical Foundations of Computing*, 2021

[10] A. Aral and T. Acar, Weighted approximation by new Bernstein-Chlodowsky-Gadjiev operators, *Filomat*, **27** (2013), 373-380.

[11] J. Boos, *Clasical and Modern Methods in Summability*, Oxford University Press, Oxford, (2000).

[12] R. C. Buck, The measure theoretic approach to density, *Amer. J. Math.*, **68** (1946), 560-580.

[13] R. C. Buck, Generalized asymptotic density, *Amer. J. Math.*, **75** (1953), 335-346.

[14] E. W. Cheney and A. Sharma, On a generalization of Bernstein polynomials, *Riv. Mat. Univ. Parma*, **2** (1964), 77-84.

[15] I. Chlodovsky, Sur le dévelopment des fonctions définies dans un intervalle infini en séries de polynomes de M. S. Bernstein, *Compositio Math.*, **4**(1937), 380-393.
[16] N. Deo and R. Pratap, α-Bernstein Kantorovich operators, *Afr. Mat.*, **31** (2020), 609-618.
[17] Z. Ditzian and V. Totik, *Moduli of Smoothness,* Springer-Verlag, New York, 1987.
[18] O. Duman and C. Orhan, Statistical approximation by positive linear operators, *Studia Math.*, **161** (2004), 187-197.
[19] J. L. Durrmeyer, Une Formule d'inversion de la Transformee Laplace. Applications a la Theorie des Moments, These de 3e Cycle, *Faculte des Science de I' Universite de Paris*, 1967.
[20] H. Fast, Sur la convergence statistique, *Coloq. Math.*, **2** (1951), 241-244.
[21] J. A. Fridy, On statistical convergence, *Analysis*, **5** (1985), 301-313.
[22] J. A. Fridy and H. I. Miller, A matrix characterization of statistical convergence, *Analysis*, **11** (1991), 59-66.
[23] V. Gupta, A note on the rate of convergence of Durrmeyer type operators for function of bounded variation, *Soochow J. Math.*, **23** (1997), 115-118.
[24] V. Gupta and R. P. Agarwal, *Convergence Estimates in Approximation Theory*, Springer, Cham, 2014.
[25] V. Gupta and W. Heping, The rate of convergence of $q-$Durrmeyer operators for $0 < q < 1$, *Math. Methods Appl. Sci.*, **31** (2008), 1946-1955.
[26] P. P. Korovkin, On convergence of linear positive operators in the space of continuous functions (Russian), *Doklady Akad. Nauk. SSSR(NS)*, **90** (1953), 961–964.
[27] M. A. Özarslan and O. Duman, Local approximation behavior of modified SMK operators, *Miskolc Math. Notes*, **11** (2010), 87-99.
[28] R. Păltănea, Durrmeyer type operators on a simplex, *Constr. Math. Anal.*, **4** (2021), 215-228.
[29] C. Prakash, N. Deo and D. K. Verma, Bézier variant of Bernstein–Durrmeyer blending-type operators, *Asian-Eur. J. Math.*, **15** (2022), Paper No. 2250103, 17 pp.
[30] C. Prakash, D. K. Verma and N. Deo, Approximation by a new sequence of operators involving Apostol-Genocchi polynomials, *Math. Slovaca*, **71** (2021), 1179-1188.
[31] R. Pratap and N. Deo, Approximation by genuine Gupta-Srivastava operators, *Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat. RACSAM*, **113** (2019), 2495-2505.
[32] K. K. Singh and P. N. Agrawal, On Szász-Durrmeyer type modification using Gould Hopper polynomials, *Mathematical Foundations of Computing*, 2022.
[33] D. Söylemez and F. Taşdelen, Approximation by Cheney-Sharma Chlodovsky operators, *Hact. J. Math. Stat.*, **49** (2020), 510-522.
[34] D. Söylemez and M. Ünver, Korovkin type theorems for Cheney-Sharma operators via summability methods, *Results Math.*, **72** (2017), 1601-1612.
[35] D. D. Stancu, The remainder in the approximation by a generalized Bernstein operator: A representation by a convex combination of second-order divided differences, *Calcolo*, **35** (1998), 53-62.
[36] D. K. Verma, V. Gupta and P. N. Agrawal, Some approximation properties of Baskakov–Durrmeyer–Stancu operators, *Appl. Math. Comput.*, **218** (2012), 6549–6556.
[37] K. Weierstrass, Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*, **2** (1885), 633-639.

Received June 2022; revised September 2022; early access September 2022.

International Journal of Environmental Science and Technology
https://doi.org/10.1007/s13762-022-04556-3

ORIGINAL PAPER

Assessment of sustainable maintenance performance of automobile garages in India

A. T. James[1] · G. Kumar[2] · A. Pundhir[2] · S. Tiwari[2] · R. Sharma[2]

Received: 26 January 2021 / Revised: 19 August 2022 / Accepted: 17 September 2022
© The Author(s) under exclusive licence to Iranian Society of Environmentalists (IRSEN) and Science and Research Branch, Islamic Azad University 2022

Abstract

Manufacturing organizations are thriving for the implementation of sustainable practices during the entire lifecycle of their products. Maintenance is an indispensable activity in the lifecycle of automobiles that consumes time, energy, money, and resources. There is considerable waste generation, energy consumption, and pollution during automobile maintenance. Hence, sustainable maintenance initiatives by garages can substantially contribute to environmental protection. However, the majority of the garages in developing countries, including India, are far behind in sustainable maintenance initiatives. This paper develops a framework for the assessment and comparison of sustainable maintenance performance of automobile garages in India based on a hybrid approach. The sustainable maintenance performance attributes are identified through literature and garage surveys. The weightage of each sustainable maintenance performance attributes is established through analytic hierarchy process. The methodology of preference ranking organization method for enrichment evaluations is used for comparing and ranking various garages according to their sustainable maintenance performance. The interrelationships among sustainable maintenance performance attributes are established through decision making trial and evaluation laboratory which segregates them into cause and effect groups. The inferences would be helpful to the garage management to initiate actions for the improvement of sustainable maintenance performance and contribute to the environment benign initiatives.

Keywords Analytic hierarchy process · Automobile maintenance · Decision making trial and evaluation laboratory · Garage · Preference ranking organization method for enrichment evaluations · Sustainable maintenance performance measurement

Introduction

All automobiles sold in India meet regulatory criteria and are reliable (James et al. 2020), but demand scheduled maintenance for long-term reliability. Automobile maintenance improves operational life, ensures safety, reduces pollution, reduces operating costs, maintains the resale value, and eliminates failures (Doshi et al., 2014). A typical vehicle repair shop cleans, serves, and repairs numerous systems, as well as provides emergency roadside assistance (Brito et al. 2007). Preventive maintenance of vehicles includes cleaning filters, replacing the engine and transmission oil, and radiator coolant, and adjusting/aligning all parts as needed (Sabharwal et al. 2010; Bouguerra et al. 2012). James and James (2020) found that the average Indian car owner visits the workshop three times a year for maintenance and repairs. With these figures and an average ownership lifetime of 8–10 years for passenger cars, a customer would visit the garage 20–25 times during the lifecycle of vehicles.

Maintenance activities generate worn-out parts, used lubricants, and cleaning intermediates that are dumped into the environment (Ajukumar and Gandhi 2013). Taking these into account, various industries across the world have recognized the significance of sustainable maintenance activities for the last decade (Ighravwe and Oke 2017a). Wang and Levrat (2014) suggested integrating ecological issues with maintenance and other researchers also advocated for the same (Henderson et al. 2014; Hennequin and Restropo 2016). In this context, automobile maintenance generates several types of waste, pollutants,

Editorial responsibility: Zhenyao Shen.

✉ G. Kumar
girish.kumar154@gmail.com

[1] A.D. Patel Institute of Technology, CVM University, Vallabh Vidyanagar, Gujarat, India

[2] Delhi Technological University, Delhi, India

Published online: 28 September 2022

and byproducts that contaminate the environment. One of the main environmental problems is the release of contaminated wastewater from garages. Every carwash uses either petrol, kerosene oil, or diesel and detergent along with water. This decreases the dissolved oxygen levels and increases the toxicity of the water. Moreover, hydrocarbons, hazardous air pollutants (HAP), and volatile organic compounds (VOCs) from underground storage tanks, vent pipes, during decantation, and gasoline filling contribute to air pollution (Edgar 2009) at garages. A study conducted by Yasin et al. (2012) on the environmental pollution from automobile garages in Pakistan revealed that each car washes use 151.4 to 227.4 L of water per vehicle. Moreover, the analysis of wastewater samples revealed that the average values of tested parameters to be 520 mg/L for biological oxygen demand (BOD), 1330 mg/L for chemical oxygen demand (COD), pH value of 8.3, Cl level of 112 mg/L, SO_4 level of 268 mg/L, iron content of 4.14 mg/L, SS content of 308.5 mg/L, total dissolved salts (TDS) level of 156.2 mg/L, phosphate content of 0.3 mg/L and oil and grease content of 1070 mg/L. The levels of BOD are 5 times higher than NEQs (National Environmental Quality Standards), COD is 7 times higher, oil and grease are 107 times more, iron is 2 times higher, and settable solids are 2 times higher. Additionally, automobile maintenance activities consume some or another form of energy, e.g., electricity is utilized to light the garage and to operate power tools used for diagnostics, disassembly, repair, cutting, welding, etc. Hence, sustainable energy consumption is a matter of concern for the garages. In addition to the above, there are physical, chemical, biological, musculoskeletal, and accidental safety hazards at automobile garages for the workers (Ahmad et al. 2016). Given India's high car population and frequent garage visits, the automotive garage industry's sustainability impact is enormous and demands attention.

Government regulations along with the financial benefits are nowadays prompting organizations to follow sustainable maintenance performance (Ighravwe and Oak 2017a). Sustainable maintenance performance (SMP) includes the execution of all necessary maintenance processes for ensuring the acceptable condition of physical assets by minimizing the negative influence of processes on the environment, making wise use of resources, and caring for the safety of all stakeholders and employees (Sari et al. 2015). However, extensive studies for understanding and analysis of sustainability challenges in maintenance are lacking (Hennequin and Restropo 2016). Moreover, Ighravwe and Oak (2017a) affirmed that maintenance professionals are deprived of a commercial tool that can help them in evaluating sustainability in maintenance practices. This underlines the necessity of a framework for the measurement of sustainability in maintenance practices at automobile garages and this research article is an attempt in this direction.

Maintenance performance measurement (MPM) is an essential function for any manufacturing sector (Muchiri et al. 2011; Pintelon and Van Wassenhove 1990). A plethora of literature is available on maintenance performance measurement (MPM) systems covering several industrial arenas (Bourne et al. 2003; Simoes et al. 2011; Kumar et al. 2013). Van Horenbeek (2014) found that most existing MPM models, techniques, and frameworks are generic and do not consider the specific industry context or goals. Hence, Swanson (2001) urged for identification of maintenance objectives or attributes and the development of evaluation schemes for it in accordance with the maintenance objectives of the concerned industry for an effective MPM. This would assist in identifying issues, developing solutions, and ultimately improving maintenance procedures (Wireman 2005).

According to the best of the knowledge of the authors, there exists no substantial work that evaluates the sustainable maintenance performance of automobile garages. Hence, there is significance in conducting research in this domain as it has got both academic as well as industrial relevance. The objective of the paper is the development of literature on attributes and sub-attributes that promote sustainable maintenance initiatives. The practical relevance of this research includes the development of a framework for sustainable maintenance performance measurement (SMPM) of garages which will be used to evaluate and compare the sustainable maintenance performance of different garages. The ultimate objective of this paper is the management of garage maintenance activities toward sustainability. In this endeavor, the sustainable maintenance performance measurement would help in rating various garages according to their initiatives toward sustainable maintenance, which will assist the garage management in identifying the weak areas of maintenance to improve upon.

The remainder of this paper is coordinated as follows. Section 2 elaborates on the attributes of sustainable maintenance of garages. An overview of the methodologies of Analytic Hierarchy Process (AHP), Preference Ranking Organization Method for Enrichment Evaluations (PROMETHE-II), and Decision Making Trial and Evaluation Laboratory (DEMATEL) for SMPM is given in Sect. 3. Section 4 applies the methodology for a case study on SMPM of 4 different vehicle brand dealer garages in India. In Sect. 5, the results are discussed, and Sect. 6 gives the concluding remarks.

Sustainable maintenance performance attributes of automobile garages

SMPM of any system demands the identification of sustainable maintenance performance attributes (SMPA) of the system. SMPA encompasses those conditions at a

workplace that enhance the sustainability of maintenance activities. However, the existing literature lacks in providing the performance attributes of maintenance activities in automobile garages. Hence, the authors have conducted a literature survey on the sustainability objectives associated with maintenance activities in various industrial domains, and attempts were made to correlate it with the context of automobile garages to single out the SMPAs. Coetzee (1998) postulated that the prime objective of maintenance function in an organization is to support the production activities ensuring the production system availability, reliability, operability, and safety with optimized cost. Kelly (1997) identified assurance of plant and environmental safety as one of the maintenance objectives. Dekker (1996) highlighted the significance of life and safety of systems, including the well-being of humans toward achieving maintenance objectives. The necessity of safety in maintenance had been highlighted by several researchers (Martorell et al. 2002; Liu 2005; Quan et al. 2007). Hennequin and Restrepo (2016) stated that ensuring the safety and health of workers during maintenance is an initiative toward sustainability. Some other researchers (Azadeh et al. 2014; Chiu and Hsieh 2016) also favored the initiatives in maintenance that take care of the well-being of technicians. Azadeh et al. (2014) developed a methodology for optimizing factors that enhances the health, safety, and environmental aspects in maintenance for sustainability. Based on the above findings, hazard control and safety can be considered one of the prime sustainability attributes in garage maintenance.

Vujanovic et al. (2012) argued that maintenance needs to be optimized for a reduction in the consumption of resources, including fuel, lubricant, spares, etc. Ighravwe and Oke (2017b) identified that maintenance policy, consumable optimization, technician training, maintenance waste reduction, salvaging, environmental impact, safety and economy are the pillars of sustainability in maintenance. Aroonsrimorakot (2010) provided an extensive list of clauses for the environmental management of garages that include framing of environment policy, its communication to employees and implementation, stocking and control of harmful substances, wastewater treatment, air pollution control, waste management, saving of electricity and water, regular upkeep of garage equipment, use of personal protective equipment (PPE), maintaining garage tidiness, emergency readiness, response, and prevention. The above findings from other domains can be easily correlated with the cleanliness of garages, their waste management, pollution control, energy and water conservation, etc.

Critical thinking over the above descriptions had directed for extensive interaction with experts in automobile garages and exploration of the available literature to explore more on the attributes that contribute toward sustainable maintenance. The outcomes in the form of identification and description of various attributes and sub-attributes that promote sustainable maintenance initiatives are described in detail in the forthcoming sub-section.

Garage management policies

The garage management and its policies and practices play a major influence in promoting sustainability. The current ISO-14040:2006 (2006) standards require all manufacturing organizations to adopt environmentally friendly activities throughout the life cycle of their products to reduce waste generation and establish recycling and remanufacturing processes. Based on these, garage management must establish and implement policies that address the environmental impact of their maintenance activities. It is preferable if the policies adopted are measurable. The first is strict compliance with all municipal pollution control regulations. However, after the garages are built, these rules are often broken for various reasons. To avoid this, management should conduct frequent internal environmental audits by a competent third party and take appropriate remedial actions to improve sustainability. All garage activities should have sustainability-focused standard operating procedures (SOPs), covering resource use, energy consumption, maintenance, waste disposal, and safety. The mere existence of SOP does not guarantee sustained maintenance. The management must inform and train all staff on how to achieve sustainability in their actions.

Garage space and ambiance

The garage space is used for a variety of purposes. This includes fault diagnosis, routine checkup, fixing and replacement of significant assemblies, oil and lubricant changes, tire services, painting, welding, upholstery fixing, testing, cleaning, stockpiling, and recovery of parts, fabrication of small components, accident recovery, and overhauls, etc. There should be adequate parking space as well. Additionally, adequate space must be earmarked in the garages for the storage of harmful substances including fuel, lubricants, coolants, refrigerants, solvents, batteries, etc. with adequate safety mandates for the prevention of its leakage and pilferage. Suitable absorbent materials including sawdust, oil absorptive fabric, etc. need to be used for preventing leakage (Aroonsrimorakot 2010). Apart from these, waste accumulation and sorting locations, counting for hazardous materials are to be explicitly allotted. Separation of floor areas into dangerous areas, emergency areas, assembly points in case of emergency, etc. with colored paint lines and provision of signage for warning as well as facilities such as fire extinguishers, first aid boxes, etc. (Motey and Coast 2013) would be beneficial. According to Dhillon (2014), most human errors and mishaps in industrial maintenance are attributed

to impoverished workplace ambiance. Hence, the garage ambiance should be amicable for laborers and clients. Dant and Bowles (2003) pointed out that the garages get dirty easily since it handles contaminating substances. The Occupational Safety and Health Council (2011) reports made it mandatory to provide a slick and clean working environment for the wellbeing and safety interests of the laborers. This is practicable just through proper cleaning of the workplace, equipment, tools, and other auxiliary supports. Apreko et al. (2015) observed that a congenial automotive garage ambiance can be created by following good housekeeping habits, providing sufficient lighting and ventilation, cleaning of floor and drainage, and provision of washrooms. Ataro et al. (2018) made a stunning disclosure that the safety in vehicle garages is constantly undermined by their poor environment. Implementation of practices like 5S (Ahmed et al. 2005) would help in keeping a decent feel in an automobile garage.

Waste handling

Different types of garage waste include solid, liquid, and gaseous waste that originate while automobile maintenance. Nowadays, a new category is being added, i.e., electronic waste (e-waste). The pollution regulatory bodies in India strictly prohibit any type of trade effluent which is any liquid waste (effluent) discharged into sewers from garages. This includes used engine and transmission oil, coolants, grease, solvents, paint, acid from a battery, fuel, degreasers, detergents, cleaning chemicals, and wastewater from a car wash (Sangodoyin 1991). These liquid wastes also constitute suspended degradable and non-bio-degradable solid waste, heavy metal ions, inorganic substances, coloring agents, etc. (Ijeoma and Achi 2011). Wastewater from garages is usually contaminated with petroleum products such as fuel, lubricants, grease, and additives (Sichinga et al. 2016). Apart from trade effluents, domestic sewage including toilets, and washbasin wastewater is also a waste. There should be water treatment facilities for trade effluents before disposal and recycling facilities for domestic sewage. Solid wastes are of different categories. Some of them are biodegradable and can be landfilled, while others can be recycled and remanufactured. Solid waste like used filters, oil seals, gaskets, gloves, hoses, containers, broken glass, etc. are not biodegradable and induce landfill problems (Ajukumar and Gandhi 2013). This necessitates dedicated disposal sites or arrangements. Batteries used oil, catalytic converters, waste rubber including tires, and plastics produced should be recycled (Chaisrisuk et al. 2019). There are several components, parts, electrical, and electronics systems, including body panels that have not exhausted their complete life cycle. These components shall be repaired and reused if it fits the use. Remanufacturing is slated to be the ideal method for reducing the solid waste disposal problem. The automobile industry is a significant sector that pioneered the remanufacturing industry (Seitz 2007). Kim et al. (2008) automobile sector has a two-third share in the global remanufacturing business. Several automobile systems, such as the engine, clutch, gearbox, alternator, and water pumps, are remanufactured on a massive scale all over the world, accounting for roughly 46 percent of all remanufactured items (Amelia et al. 2009). Hence, garages must resort to remanufacturing.

Pollution control

It is observed that automobile garages create lots of environmental pollution thanks to undisciplined work practices (Ishola et al. 2017). Demie (2015) pointed out that automobile garages are one of the major sources of hazardous pollution that contaminates soil and water and Pathberiya et al. (2017) also raised the same concerns. Air pollution due to various garage activities is a matter of concern. The pollution control authority norms in India mandated that there should not be any kind of fugitive emissions that pollute air from garages. One of the sources of fugitive emission is spray painting. The pollution control norms mandate a paint booth for all spray-painting activities. Hence, garages must have a provision for a paint booth. Tian and Wang (2020) argued for green painting process innovations and implementations in garages that include the utilization of eco-friendly paint materials, reduction of waste, process optimization, etc. Care should be taken during the maintenance of car air-conditioning systems as the leakage of refrigerants is a very harmful atmosphere. This can be controlled using an air-conditioner gas recovery machine. Moreover, leakages from gas welding units must be prevented. This is possible with the use of gas leak detectors. Another one causing air pollution is the tiny particles emanating from the maintenance of different systems that pollute the air. OSHA (Occupational Safety and Health Standards) standards recommend NPEU (negative pressure enclosure unit) to alleviate the exposure of technicians to carcinogenic particles (James et al. 2017). Moreover, OEM recommends a HEPA (high-efficiency particulate arrestance) system to repress pollution due to invasion of foreign matters during the repair of systems like hydraulics and engine. The Indian pollution control norms for garages mandate that the suspended particulate matter in garage premises should not be beyond 100 µg/cubic meter. Garages must adhere to these norms. Sound pollution is another concern in garages. The statutory norms in India restrict the noise level maximum of up to 80 decibels within the premises of the garage and ban all noise-producing activities during nighttime. The techniques mentioned in the previous sub-section for the systematic handling of waste can reduce soil and water pollution.

Energy and water conservation

The energy use associated with automobile maintenance includes the consumption of different forms of energy for the maintenance processes, and general-purpose usage, including its abuse and misuse. The employees must be educated about the judicious use of electricity for illumination, climate control, etc. The workshop layout should be energy efficient that facilitate the use of natural light for illumination, reduction of movements, and minimization of HVAC (Heating Ventilating and Air Conditioning) requirements. The lack of maintenance of garage equipment and tools makes it more energy-consuming. Hence, proper maintenance and upkeep of support equipment and tools is a must. The use of gravity for material handling and related activities will help to reduce energy consumption. The utilization of solar energy or any other renewable energy source for electricity generation would be an added initiative toward energy self-sustainability. The energy audit will help in identifying machines and other auxiliary equipment like a pump, air conditioners, motors, fans, etc. that are no more energy efficient because of age, wear, etc. These machines could be replaced by new energy-efficient machines. Water usage is enormous in garages for car washing and other domestic uses and one should be judicious in its usage. Water recycling as a part of waste management practices directly conserves water. Since garages have a lot of roof areas, practices like rainwater harvesting can significantly add to it.

Hazard control and safety

There are different types of hazards associated with garage maintenance that affect the safety of maintenance technicians. The safety risks in maintenance had been exclusively discussed by several researchers (Reason 1997; Reason and Hobbs 2003). Lind and Neonen (2008) surveyed to find out what causes safety-related accidents in different industries. They found that the absence of well-defined procedures, inadequate provision of PPE, lack of technician knowledge or experience, poor working ambiance, lacunae in task planning, and inadequate safety management are major culprits. This mandates the development of a well-defined safety plan and procedure and its execution in the garage. Compliance with applicable OSHA laws and regulations would help in this regard. However, Cruickshank (2010) cautioned that garages should not underestimate occupational safety requirements and must be fastidiously overseen and controlled. This is to be ensured through third-party safety audits. Additionally, all the employees must be trained for following safety protocols, including first respondent training. Mock drills also should be a part of training. The safety instructions must be displayed at all prominent locations in the garage. There are possibilities of technicians getting injured by keeping in contact with sharp edges of components, parts, fasteners, and vehicle bodies during service and maintenance. Additionally, implications of occupational hazards on garage technicians need to be diligently addressed by garage organizations. The occupational hazard in garages is attributed to thermal and ultraviolet radiation exposures, hot and noisy workplaces, wrong tools, arduous work stances (Vyas et al. 2011), and ill psychosocial work environments (Hunt et al. 2000; ILO 2000; Rongo et al. 2004). Moreover, garage technicians come in contact with heavy metals, chemical reagents, washing detergents, rust cleaners, lubricant oils, degreasing agents, solvents, battery acid, welding spark, and fumes, spray paint particles, asbestos from tribology elements, and car exhaust gas. (Schwartz 1987; El-Saadawy et al. 2011). Prolonged exposure to the above substances is harmful to health. The garage must be free from hazards arising out of fire, electricity, chemicals, and elevated automobiles on the hoist. The maintenance waste including used batteries, brake liner materials antifreeze solutions, lubricants, catalytic converters, absorbents, etc. needs to be systematically segregated and disposed. Research regarding the occupational health hazards associated with automobile mechanics in several countries revealed that automobile maintenance technicians suffer from health complications including injury, pain, respiratory problems, skin diseases, cancers, and eye problems, (Muttamara and Alwis 1994; Kant et al. 1990). Adela et al. (2012) pointed out that prolonged exposure to the lead content in garage waste can cause fatigue as well as abdominal and central nervous dysfunction in workers. Adherence to safety rules and the mandated use of PPE can help to prevent occupational hazards to a considerable extent. This may include the use of elbow wraps, hand gloves, apron, safety shoes, helmets, face shields, safety goggles, earplugs, etc. (Parker et al. 2012; James et al. 2017). There fire extinguishers and first aid facilities should be mandatory for automobile garages. Motey and Coast (2013) have presented an extensive and exclusive list of safety precautions for automobile garages. Some of them are listed here. All the machines should have a protective guard and all electrical connections must be properly insulated. Wearing PPE for any garage activity should be a habit. The technicians must ensure that their hands are clean, clear, and dry while using any hand tool to avoid slipping. Technicians must keep away their hands and loose clothes away from power tools and ensure that the safeguards of power tools are closed before their operation. They must turn off the tools once the process is over. Care must be taken while using the compressor and check all its connections and nozzles. The usage of a hydraulic jack or hoist for lifting the vehicle should be with utmost care. There should not be any occupant in and beneath

the vehicle, while it is being lifted. Additional jack stands or supports can be used for extra protection. Flammable substances should be kept in fireproof boxes and the entire garage should be a no-smoking zone. Care should be taken while working with hot parts such as engines, and radiators, as well as poisonous and inflammable substances like coolants, battery acid, refrigerants, etc. Extreme care must be taken while repairing electrical systems and removing or disconnecting the ground terminal of the battery to prevent electrocution. Extra care is to be made while doing the test drives. No one should stand in front or behind the vehicle during the test run. Garages need to prepare a checklist based on these guidelines and follow it before any operations.

Table 1 summarizes and represents all of the above-mentioned sustainable maintenance attributes and sub-attributes.

Materials and methods

The sustainable maintenance performance evaluation and rating of garages form a complex evaluation system. Moreover, a causal relationship among SMPAs is also required to be established. Hence, a hybrid methodology of AHP, preference ranking organization method for enrichment evaluations (PROMETHEE-II), and DEMATEL is suggested for the modeling. The approaches are described in detail in the following subsections.

Table 1 Attributes of sustainable garage maintenance

Attribute	Sub-attributes	Notation
Garage management policies (GMP)	Compliance with pollution control norms	GMP_1
	Periodic internal environment audit	GMP_2
	Existence of SOP for sustainable practices	GMP_2
	Communication and training on environmental norms	GMP_4
Garage space and ambiance (GSA)	Space provision for repair, service, and parking	GSA_1
	Space provision for storage of harmful substances	GSA_2
	Space provision for waste accumulation and sorting	GSA_3
	Marking and signage	GSA_4
	Garage tidiness	GSA_5
	Garage lighting	GSA_6
	Practice of 5S	GSA_7
Waste handling (WH)	Water treatment facilities	WH_1
	Water recycling facilities	WH_2
	Dedicated waste disposal site availability	WH_3
	Solid and liquid waste recycling practices	WH_4
	Solid waste repair/remanufacturing practices	WH_5
Pollution control(PC)	Provision of paint booth	PC_1
	Provision of AC gas recovery machine	PC_2
	Provision of gas leak detector	PC_3
	Provision of NPEU and HEPA	PC_4
	Conformance to noise pollution norms	PC_5
Energy and water conservation(EWC)	Maintenance and upkeep of garage equipment	EWC_1
	Renewable energy usage	EWC_2
	Energy auditing	EWC_3
	Rainwater harvesting	EWC_4
Hazard control and safety(HCS)	The practice of OSHA laws	HCS_1
	Safety audit	HCS_2
	Safety training	HCS_3
	Display of safety instructions	HCS_4
	Safety check listing practices	HCS_5
	Personal protective equipment (PPE) provision	HCS_6
	First aid facility	HCS_7
	Fire extinguisher provision	HCS_8

Analytic hierarchy process method

Analytic Hierarchy Process is a dynamic problem-solving tool for complicated and fuzzy situations that cannot be quantified. It is commonly used in subjective decision-making processes in a hierarchical system based on multiple criteria (Tzeng and Huang 2011). According to Saaty (2008), AHP is based on the measurement theory of pairwise comparisons and is built on the subjective judgments of domain experts to establish priority scales. The AHP paradigm aids in issue conceptualization and structure (Chan 2006). The AHP approach had been effectively utilized for analyzing environmental impact (Ramanathan 2001), optimizing sustainable supply chains (Resat and Unsal 2019), and assessing urban vulnerability in the Amazon Delta and Estuary (Mansur et al. 2016), etc. In this work, AHP is used to evaluate the weightages of SMPAs.

Procedure of AHP

The sequence of steps for applying AHP (Saaty 1994) for assessing the weightages is consolidated as:

Step 1: Development of pairwise comparison matrix.

Considering there are 'N' elements in the system to be evaluated, then the pairwise comparison of element 'i' with element 'j' will yield a square '$N \times N$' matrix 'S_1'. In this matrix, each element e_{ij} depicts the relative importance of element 'i' with respect to 'j'. In the matrix 'S_1', $e_{ij} = 1$ when $i = j$ and $e_{ji} = 1/e_{ij}$. For pairwise comparison of elements in the system, Saaty's 1 to 9 scale (Saaty, 1994) mentioned in Table 2 is utilized.

Step 2: Calculation of normalized weight and maximum eigenvalue

The pairwise comparison matrix is normalized by dividing each entry in a column by the sum of column entries of 'S_1'.

Table 2 Allocating value to pairwise comparisons of matrix elements (Saaty 1994)

Verbal judgment of relative importance between elements 'i' and 'j'	Numerical rating
Equal importance	1
Moderate importance of one element over other	3
Strong importance	5
Very strong importance	7
Extremely strong importance	9

Let 'b_{ij}' be the normalized value of ith element with respect to jth element and it is expressed as:

$$b_{ij} = \frac{a_{ij}}{\sum_{i=1}^{n} a_{ij}} \text{ For all } j = 1 \text{ to } n. \quad (1)$$

The normalized matrix is used to calculate the normalized relative weightage of matrix elements by averaging the values in each row. Using Eq. (2), compute the normalized weight (WE_i) of each element.

$$WE_i = \frac{\sum_{j=1}^{n} b_{ij}}{n} \quad (2)$$

Create matrices S_3 and S_4 such that $S_3 = S_1 x S_2$ and $S_4 = S_3 / S_2$, where $S_2 = [WE_i]^T$ where 'T' is the matrix transpose. Evaluate the maximum eigenvalue 'λ_{max}' which is the average of matrix S_4.

Step 3: Establish the Consistency Index (CI).

The consistency index is calculated as follows $CI = (\lambda_{max} - N)/(N-1)$. A smaller CI value indicates a little consistency departure.

Step 4: Establishment of the Random Index (RI).

With reference to Table 3, calculate the random index (RI) for the system elements.

Step 5: Determine the consistency ratio (CR).

Determine the consistency ratio as CR = CI/RI. In general, a consistency ratio (CR) of 0.1 or less is considered fair, indicating a reasonable decision-maker viewpoint in allocating the relative importance/weightage of factors.

PROMETHEE methodology

Preference ranking organization method for enrichment evaluation (PROMETHEE) was introduced in 1982 by J.P. Barns (1982). Among the several versions of PROMETHEE available in the literature, this paper makes use of the PROMETHEE-II method (Brans and Vincke 1985; Brans et al. 1986) because it concludes the full ranking of the evaluated factors which is required in this work. This methodology has been successfully applied for the selection and prioritization of green area irrigation resources (Safari et al. 2020), risk prediction of leaks in storage tanks (Ikwan et al. 2020), and environmental performance evaluation of suppliers (Tuzkaya et al. 2009), etc. In this work, the PROMETHEE-II method is applied for assessing the sustainable maintenance performance of the garages and their ranking. The following are the steps suggested for the implementation of the PROMETHEE-II methodology.

Table 3 Values of random index (Saaty 1994)

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.85	0.9	1.12	1.24	1.32	1.41	1.45	1.51

Procedure of PROMETHEE

Step 1: Normalization of the decision matrix (D).

After taking the inputs from the experts, the decision matrix (D) is formulated, which is used as the initial matrix for the PROMETHEE. Subsequently, matrix 'D' is normalized to obtain normalization matrix (R) by using expression (3)

$$R_{\mathrm{ab}} = \frac{D_{\mathrm{ab}} - \min(D_{\mathrm{ab}})}{\max\left(D_{\mathrm{ab}}\right) - \min(D_{\mathrm{ab}})} \tag{3}$$

Step 2: Determination of the deviation by pairwise comparison.

Normalized matrix (R) is further evaluated for getting the pairwise deviation by subtracting the value of each criterion from that of the other according to expression (4)

$$\mathrm{db}(x, y) = \mathrm{rb}(x) - \mathrm{rb}(y) \tag{4}$$

where $d_b(x,y)$ is the difference between the value 'x' and 'y' on each criterion and 'r_b' is the value of the Normalization matrix (R).

Step 3: Calculate the preference function.

A preference function is obtained based on expression (5)

$$\mathrm{Pb}(x, y) = Fd[\mathrm{db}(x, y)] \tag{5}$$

where, $F(d)=0$, if $d<0$ and $F(d)=\mathrm{d}$, if $d>0$, i.e., if deviation of pairwise comparison is negative, the function 'F' deduce its value to 0 and if obtained deviation is positive, the function 'F' will retain its actual value in matrix $P(x,y)$.

Step 4: Calculate the aggregated preferences.

Aggregated preferences or Global preferences are calculated by multiplying the criteria weight (W_b) that is obtained from the AHP method with the values of $P(x,y)$.

$$\sigma(x, y) = \sum_{b=1}^{k} P(x, y).\mathrm{Wb} \tag{6}$$

Step 5: Determination of the outranking flows/partial ranking.

Leaving outranking flow ($\phi+$) and Entering outranking flow ($\phi-$) are calculated by taking the mean values of the rows and columns, respectively, of matrix $\sigma(x, y)$. They are used to denote the partial ranking by the PROMETHEE-I method. However, PROMETHEE-II involves a step more to obtain the complete ranking of the alternatives.

Leaving outranking flow (ϕ^+) is determined by

$$\phi^+ = \frac{1}{n-1} \sum_{c \in \mathrm{A}} \sigma(x, c) \tag{7}$$

Entering outranking flow (ϕ^-) is determined by

$$\phi^- = \frac{1}{n-1} \sum_{c \in \mathrm{A}} \sigma(c, x) \tag{8}$$

Step 6: Obtaining the complete ranking.

Obtain the net outranking flow $\phi(x)$ by subtracting 'Entering outranking flow (ϕ^-)' from 'Leaving outranking flow (ϕ^+) as per Eq. (9).

$$\phi(x) = \phi^+ - \phi^- \tag{9}$$

The net outranking flow $\phi(x)$ determines the final ranking of the alternatives, Higher the value of the $\phi(x)$ of an alternative, the greater will be its ranking in the net ranking.

DEMATEL method

Decision making trial and evaluation laboratory (DEMATEL) method is regarded as an efficient method for identifying interdependencies among several factors in a complex system. (Sheng-Li et al. 2018). Geneva Research Centre developed the DEMATEL method between 1972 and 1976 for effectively categorizing factors into cause and effect groups. Additionally, the mathematical model also assists in recognizing the critical factors through an impact diagram. Unlike the conventional quantitative tools such as AHP, which follow the assumption of all factors being independent, DEMATEL considers all relations among the factors through contextual relationships. DEMATEL has been used effectively in the analysis of the challenges for environmentally sustainable manufacturing in the Indian automobile industry (Mathiyazhagan et al. 2018), the analysis of factors hindering the implementation of e-waste management strategies (Kumar and Dixit 2018), the analysis of barriers to remanufacturing (Bhatia and Srivastava 2018), and other areas. The DEMATEL method is utilized in this paper to develop relationships among SMPAs of automobile garages.

Procedure of DEMATEL

Step 1: Establish an initial direct relation matrix (M).

A pairwise comparison matrix among attributes is obtained through the rating provided by each expert. To compare the impact between two attributes, a 0–5-point scale is used as per Table 4.

Multiple experts are submitting their views, and it is critical to consider all respondents' advice. For example, if 'z' is the number of experts and 'p' is the number of characteristics with $1 \le z \le Y$, then there will be 'p × p' non-negative matrices for each expert. The entry 'x_{mn}' refers to the extent to which factor 'm' influences factor 'n'. Now, to include all responses from 'Y' experts, an average matrix 'a_{mn}' is constructed using Eq. (10)

Table 4 Scale for pairwise comparison

Definition	Numerical rating
No influence	0
Very low influence	1
Low influence	2
Medium influence	3
High influence	4
Very high influence	5

$$a_{mn} = \frac{1}{Y}\sum_{p=1}^{Y} \mathrm{x}_{\mathrm{mn}}^{\mathrm{p}} \tag{10}$$

Step 2: Development of the direct relation matrix (B). The direct relation matrix is obtained by Eq. (11)

$$B = A \times S \tag{11}$$

$$\text{where, } S = \min\left[\frac{1}{\max\sum_{n=1}^{p}|m_{\mathrm{mn}}|}, \frac{1}{\max\sum_{n=1}^{p}|m_{\mathrm{mn}}|}\right] \tag{12}$$

Step 3: Establish the total relation matrix (C). The total relation matrix (C) is given by Eq. (13)

$$C = B(I - B)^{-1} \tag{13}$$

The total relation matrix is created by adding the sums of all the rows and columns. The element 'Y_n' reflects both direct and indirect effects imparted by other attributes on factor 'n,' whereas the element 'X_m' represents all effects received by other attributes from the attribute 'm.' Consider the value of $(X_m + Y_n)$ to determine the prominence of an attribute. However, the difference between the two values 'X_m' and 'Y_n' is evaluated to determine whether an attribute belongs to the cause or effect category.

Step 4: Evaluate the threshold value.

The threshold value is determined by averaging all of the components in the total relation matrix (C). It can also be used to create a cause-and-effect diagram. The threshold value serves as a critical value because only values with magnitudes greater than the threshold value are depicted in the cause-effect diagram. The produced causal diagram shows interaction links between the qualities and aids in determining the interdependencies and behavior of the attributes among themselves.

Application of the methodology

In this section, the methodology is applied for the sustainable maintenance performance evaluation of automobile garages. For this, four different dealer garages corresponding to four different major brands of passenger cars were randomly chosen. The garages were situated in one of the municipal corporations of western India. All the garages were of high repute, offering a similar quality of service and handling nearly the same number of vehicles daily. To maintain anonymity, the garages were designated as G_1, G_2, G_3, and G_4. For the assessment and comparison of garages, a team of experts including academicians, maintenance engineers, environment engineers, and safety professionals is constituted. The following procedures are undertaken for the application of the methodology.

Table 5 Normalized relative weights of sustainable maintenance attributes and ranking for importance

Sr. no.	Sustainable maintenance attributes	Weightage (W_b)	Rank
1	GMP	0.032	6
2	GSA	0.061	5
3	WH	0.14	3
4	PC	0.242	2
5	EWC	0.068	4
6	HCS	0.457	1

Assessment of weightages of sustainable maintenance attributes

The first step toward performance measuring is identifying the weightage of each sustainable maintenance attribute. This is accomplished by following the Technique actions outlined in Sect. 3.1.1. First, a pairwise comparison square matrix of sustainability attributes using Saaty's 1–9 scale is created based on the recommendations of a team of experts. This is represented as matrix 'S_1' in expression (10).

$$S_1 = \begin{array}{c} \text{Sustainable maintenance attributes} \\ \text{GMP} \\ \text{GSA} \\ \text{WH} \\ \text{PC} \\ \text{EWC} \\ \text{HCS} \end{array} \begin{array}{c} \begin{array}{cccccc} \text{GMP} & \text{GSA} & \text{WH} & \text{PC} & \text{EWC} & \text{HCS} \end{array} \\ \begin{bmatrix} 1 & 0.33 & 0.2 & 0.14 & 0.5 & 0.11 \\ 3 & 1 & 0.5 & 0.25 & 0.5 & 0.12 \\ 5 & 2 & 1 & 0.5 & 3 & 0.25 \\ 7 & 4 & 2 & 1 & 4 & 0.5 \\ 2 & 2 & 0.33 & 0.25 & 1 & 0.12 \\ 9 & 8 & 4 & 2 & 8 & 1 \end{bmatrix} \end{array} \tag{14}$$

The normalized weights of sustainable maintenance objectives are established by systematically following the processes described in Sect. 3.1.1 for AHP, and they are ranked and shown in Table 5 as a result. The consistency ratio (CR) of the pairwise comparison matrix is then calculated to see if the expert responses are consistent. Expert

responses are considered consistent if the CR value is less than 0.1. If the CR value is more than 0.1, the nature of the judgments should be altered until the CR value falls within a consistent range. The consistency ratio is calculated using the procedure outlined in Sect. 3.1.1. In this case study, the CR value achieved is 0.0295, which is less than 0.1, indicating a high level of consistency in expert opinion.

Application of PROMETHEE-II for performance evaluation and ranking of garages

Once the relative weights of maintenance attributes are established, the MCDM (Multi-Criteria Decision Making) method named PROMETHEE-II is used for ranking the garages considering the weights of all attributes. To initiate this, the sustainable maintenance sub-attributes for each attribute were evaluated on a scale of 0 to 2 by the team of experts after the visit to each of the garages. A discrete value is assigned on levels, e.g., None (N), Average (A), and High (H), with a value of 0, 1, and 2, respectively, for sub-attributes. There is a probability that subjectivity is being cropped up in such an assignment of a score. However, since the task is done by a team of experts, it gets minimized. Tables 6, 7, 8, 9, 10 and 11 are developed by the experts as a guideline for the assessment of the other five SMPAs such as GMP, GSA, GA, WH, PC, and HCS, respectively.

The experts have evaluated different garages as per the guidelines in Tables 6, 7, 8, 9, 10 and 11. The observations in the linguistic form are displayed in Table 12.

The attainment score of sustainable maintenance attributes for the garages in numerical form is given in Table 13. This essentially forms the decision matrix (D) which is mentioned in Sect. 3.2.1.

Using Eq. (3), the normalized decision matrix (R) is prepared and shown in Table 14.

Table 6 Guidelines for the assessment of sub-attributes of garage management policies (GMP)

Sub-attributes	None (N)	Average (A)	High (H)
Compliance with pollution control norms	Minimum compliance	Major compliance, but lacuna in regular upkeep	Complete compliance with norms and their upkeep
Periodic environment audit	No audit	Inhouse environment audit	External audit
Existence of SOP for sustainable practices	No SOP	SOP for selected processes	SOP for all processes
Communication and training on environmental norms	No communication and training	Inhouse training and communication through supervisors	Induction training for all employees by an external expert and periodic communication

Table 7 Guidelines for the assessment of sub-attributes of garage space and ambiance (GSA)

Sub-attributes	None (N)	Average (A)	High (H)
Space provision for repair, service, and parking	Inadequate space for task manipulation and storing of spare parts, fasteners, etc	Enough space for task manipulation and storing	Ergonomically construed workplace with space earmarked for designated activities
Space provision for storage of harmful substances	No separate storage space	Average provision	Dedicated space with protective measures and absorbent material
Space provision for waste accumulation and sorting	No separate space	Space provided, but inadequate prevention for the spread of waste	Dedicated space with prevention of the spread of waste suggested by competent authorities
Marking and signage	No marking and signs	Marking and signage of critical areas	Marking and signage for all areas
Garage tidiness	Incredibly poor cleanliness of floors, devices, store and different territories of garage	The average level of neatness	Amazingly perfect and clean garage, with 5S standards implemented in place
Garage lighting	Inadequate lighting	Illuminated with normal lights, including sunlight during daytime. Additionally, torchlights are given for typical applications and emergency use	Exceptionally radiant bright spotlights are provided in the garage
Practice of 5S	No implementation of 5S practice	Partial follow-up of 5S practices	5S practices are carefully followed with endeavors for ceaseless improvement

Table 8 Guidelines for the assessment of sub-attributes of waste handling (WH)

Sub-attributes	None (N)	Average(A)	High(H)
Water treatment facilities	No treatment	Partial treatment	Complete treatment
Water recycling facilities	No recycling	Partial recycling	Complete recycling
Dedicated waste disposal site availability	No dedicated site	Partial provision	Dedicated site with all protective measures and absorbent material
Solid and liquid waste recycling practices	No recycling	Only critical wastes are recycled through the legal waste management company	The dedicated in-house and legal waste management company for recycling the majority of waste
Solid waste repair/remanufacturing practices	No repair /remanufacturing	Less than 20% of components are repaired/remanufactured	Up to 40% of components are repaired/remanufactured

Table 9 Guidelines for the assessment of sub-attributes of pollution control (PC)

Sub-attributes	None (N)	Average (A)	High (H)
Provision of paint booth	No paint booth	In-house developed paint booth	Standard paint booth as per norms
Provision of AC gas recovery machine	No provision	Brought from outside based on demand	Inhouse facility
Provision of gas leak detector	No provision	Brought from outside based on demand	Inhouse facility
Provision of NPEU and HEPA	No provision	Partial provision	Complete provision of NPEU and HEPA
Conformance to noise pollution norms	No conformance	Occasional violation of norms	Complete conformance

Table 10 Guidelines for the assessment of sub-attributes of energy and water conservation (EWC)

Sub-attributes	None (N)	Average (A)	High (H)
Maintenance and upkeep of garage equipment	Breakdown maintenance policy for garage equipment	Condition-based maintenance	Dedicated preventive maintenance scheduling
Renewable energy usage	No usage	Less than 50% of the energy used is renewable	More than 50% of the energy used is renewable
Energy auditing	No energy audit	In house audit	Third-party external audit
Rainwater harvesting	No practice	25% of water requirement is harvested	More than 25% of the water requirement is harvested

Table 11 Guidelines for the assessment of sub-attributes of hazard control and safety (HCS)

Sub-attributes	None (N)	Average (A)	High (H)
The practice of OSHA laws	No practices	Average adherence	Strict adherence
Safety audit	No audit	Once in year	Once in 6 months
Safety training	No training	In house training	Training by external experts
Display of safety instructions	Only a few places	Average display	Adequate display
Safety check listing practices	No practices	Checklist for critical activities only	Checklist for all activities
Personal protective equipment (PPE) provision	Provision of only safety shoes	Complete arrangement of all PPE and safety gadgets for usage	Complete provision of all PPE and safety gadgets with enforced usage
First aid facility	No first aid facility	Partial provision	Very good facility as per norms
Fire extinguisher provision	No fire extinguisher	Dedicated water line, form extinguisher, and fire buckets are available	Dedicated water line along with dry chemical fire extinguishers and CO_2 extinguishers for dousing A, B, C, and D types of fires

Table 12 Expert evaluation results of sustainable maintenance performance of garages

Sub-attributes	Garages			
	G_1	G_2	G_3	G_4
GMP_1	A	A	H	H
GMP_2	A	A	A	H
GMP_3	A	H	H	A
GMP_4	A	A	A	A
GSA_1	A	H	H	A
GSA_2	H	H	A	H
GSA_3	A	A	A	A
GSA_4	H	A	H	H
GSA_5	A	A	H	H
GSA_6	H	A	A	H
GSA_7	A	A	A	H
WH_1	N	N	N	A
WH_2	N	N	N	A
WH_3	A	A	A	H
WH_4	N	N	A	A
WH_5	A	N	N	N
PC_1	H	H	H	H
PC_2	H	H	H	H
PC_3	N	H	H	H
PC_4	N	N	A	A
PC_5	A	A	A	A
EWC_1	H	H	A	H
EWC_2	A	N	N	A
EWC_3	A	N	N	A
EWC_4	N	A	A	H
HCS_1	A	A	H	A
HCS_2	A	H	H	A
HCS_3	H	A	A	A
HCS_4	A	H	H	A
HCS_5	A	A	A	A
HCS_6	A	A	A	H
HCS_7	A	A	A	A
HCS_8	A	A	A	H

Table 13 The decision matrix (*D*) of attainment score of sustainable maintenance attributes

Sustainable maintenance attribute	Garages			
	G_1	G_2	G_3	G_4
GMP	4	5	6	6
GSA	10	9	10	12
WH	2	1	2	5
PC	5	7	8	8
EWC	4	3	2	6
HCS	9	10	11	10

The pairwise comparison matrix of SMPA for PROMETHEE is developed using Eq. (4). The pairwise matrix is shown in Table 15. This will help in identifying the difference between respective values of attributes for every possible pair of garages. For instance, G_{23} shows the difference between respective attribute values for G_2 and G_3.

The multi-criteria preference index matrix is then developed using Eq. (6) and presented in Table 16.

To calculate the outranking flows, the positive flow and negative flow is calculated using Eqs. (7) and (8). The difference between positive flow and negative flow is calculated to obtain the value of net flow using Eq. (9) and shown in Table 17. Based on the value of net flow, the ranking of garages is obtained.

Application of DEMATEL for the causal relationships among attributes

DEMATEL method is applied for determining the interdependencies among the listed attributes. It helps in the evaluation of the inter-relationship among the sustainable maintenance performance attributes by constructing a cause-effect diagram. After identifying the attributes and listing them, each expert is asked to prioritize the attributes as per Table 4. By taking the average of the responses of all the experts and using Eq. (10), the average matrix (a_{mn}) is obtained and shown in Table 18.

Using Eq. (11), the direct relation matrix (B) of the attributes is developed and shown in Table 19.

The total relation matrix 'C' is developed by using Eq. (13) and shown in Table 20. In Table 20, the $(X_m + Y_n)$ column shows the degree of prominence of each attribute over the entire system. The higher the value of $(X_m + Y_n)$, the greater the significance of that attribute. However, the $(X_m - Y_n)$ value determines whether an attribute belongs to the cause group or effect group. If $(X_m - Y_n) < 0$ for any attribute, then it belongs to the effect group i.e., "being influenced by other" and if $(X_m - Y_n) > 0$, then the attribute belongs to the cause group i.e., "influencing other".

Results and discussion

AHP was employed to analyze the significance of each attribute in terms of weightage. The weightages obtained through AHP were further used in PROMETHEE-II for ranking the garages in terms of sustainable maintenance performance. Subsequently, the DEMATEL method was implemented to identify the most prominent sustainable attribute among the SMPAs. Also, the attributes were divided into cause and effect groups. The prioritization and classification of SMPAs will provide a better understanding for the

Table 14 Normalized decision matrix (R)

	GMP	GSA	WH	PC	EWC	HCS
G_1	0	0.333333	0.25	0	0.5	0
G_2	0.5	0	0	0.666667	0.25	0.5
G_3	1	0.333333	0.25	1	0	1
G_4	1	1	1	1	1	0.5

Table 15 Pairwise comparison matrix of SMPA for PROMETHEE

G_{12}	−0.5	0.333333	0.25	−0.66667	0.25	−0.5
G_{13}	−1	0	0	−1	0.5	−1
G_{14}	−1	−0.66667	−0.75	−1	−0.5	−0.5
G_{21}	0.5	−0.33333	−0.25	0.666667	−0.25	0.5
G_{23}	−0.5	−0.33333	−0.25	−0.33333	0.25	−0.5
G_{24}	−0.5	−1	−1	−0.33333	−0.75	0
G_{31}	1	0	0	1	−0.5	1
G_{32}	0.5	0.333333	0.25	0.333333	−0.25	0.5
G_{34}	0	−0.66667	−0.75	0	−1	0.5
G_{41}	1	0.666667	0.75	1	0.5	0.5
G_{42}	0.5	1	1	0.333333	0.75	0
G_{43}	0	0.666667	0.75	0	1	−0.5

Table 16 Multi-criteria preference index matrix for PROMETHEE

	Weight of SMPA (W_b)						SUM
	0.032	0.061	0.14	0.242	0.068	0.457	
G_{12}	0	0.020333	0.035	0	0.017	0	0.072333
G_{13}	0	0	0	0	0.034	0	0.034
G_{14}	0	0	0	0	0	0	0
G_{21}	0.016	0	0	0.161333	0	0.2285	0.405833
G_{23}	0	0	0	0	0.017	0	0.017
G_{24}	0	0	0	0	0	0	0
G_{31}	0.032	0	0	0.242	0	0.457	0.731
G_{32}	0.016	0.020333	0.035	0.080667	0	0.2285	0.3805
G_{34}	0	0	0	0	0	0.2285	0.2285
G_{41}	0.032	0.040667	0.105	0.242	0.034	0.2285	0.682167
G_{42}	0.016	0.061	0.14	0.080667	0.051	0	0.348667
G_{43}	0	0.040667	0.105	0	0.068	0	0.213667

Table 17 Net flow matrix for PROMETHEE

Garages	Positive flow	Negative flow	Net flow	Rank
G_1	0.035444	0.606333	−0.57089	4
G_2	0.140944	0.267167	−0.12622	3
G_3	0.446667	0.088222	0.358444	1
G_4	0.414833	0.076167	0.338667	2

decision-makers in terms of the identification of attributes that needs attention for enhanced performance.

Based on AHP analysis, HCS and PC were the most significant attributes with weightages of 0.457 and 0.242, respectively. The weightage values obtained were used further in PROMETHEE-II analysis for ranking of the garages. Based on the results obtained from PROMETHEE-II analysis, garages were arranged according

Table 18 Average initial direct-relation matrix (a_{mn}) for DEMATEL

Attributes	GMP	GSA	WH	PC	EWC	HCS	SUM
GMP	0.000	2.000	1.667	2.333	2.333	2.333	10.667
GSA	1.000	0.000	3.000	2.667	3.333	3.000	13.000
WH	2.000	2.667	0.000	3.333	2.333	2.667	13.000
PC	3.333	2.667	3.333	0.000	1.667	2.667	13.667
EWC	3.667	2.667	1.667	2.333	0.000	2.667	13.000
HCS	2.667	2.333	3.667	3.000	3.000	0.000	14.667
							14.667

Table 19 Direct relation matrix (B) for DEMATEL

Attributes	GMP	GSA	WH	PC	EWC	HCS
GMP	0	0.136	0.114	0.159	0.159	0.159
GSA	0.068	0	0.205	0.182	0.227	0.205
WH	0.136	0.182	0	0.227	0.159	0.182
PC	0.227	0.182	0.227	0	0.114	0.182
EWC	0.250	0.182	0.114	0.159	0	0.182
HCS	0.182	0.159	0.250	0.205	0.205	0

Table 20 Total relation and direct–indirect influence matrix for DEMATEL

Attributes	GMP	GSA	WH	PC	EWC	HCS	X_m	X_m+Y_n	X_m-Y_n
GMP	0.986	1.079	1.132	1.184	1.113	1.160	6.653	14.386	−1.080
GSA	1.258	1.157	1.412	1.417	1.363	1.403	8.010	15.548	0.471
WH	1.295	1.299	1.231	1.437	1.301	1.375	7.938	16.022	−0.147
PC	1.391	1.331	1.452	1.289	1.304	1.411	8.179	16.415	−0.058
EWC	1.357	1.275	1.306	1.362	1.145	1.352	7.797	15.472	0.123
HCS	1.447	1.397	1.552	1.547	1.450	1.343	8.736	16.780	0.692
Y_n	7.733	7.538	8.085	8.237	7.675	8.044	Threshold value = 1.314		

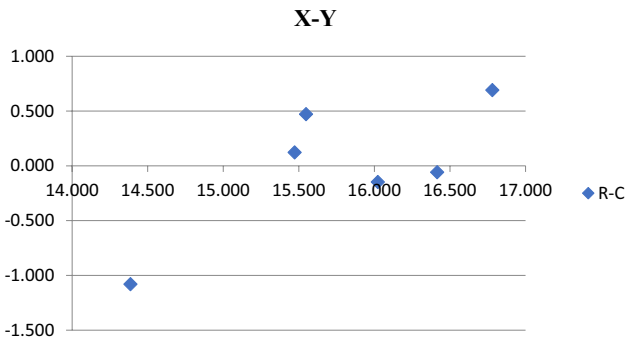


Fig. 1 Cause-effect diagram of SMPA for categorization

to their maintenance performance in the order G_3, G_4, G_2, and G_1.

Using DEMATEL methodology, it was found that HCS and PC are the most prominent attributes with the highest 'X_m+Y_n' value, and GMP and EWC are the least prominent attributes with the lowest 'X_m+Y_n' value. A cause-effect diagram is drawn by considering (X_m+Y_n) along the x-axis and (X_m-Y_n) along the y-axis and shown in Fig. 1. The diagram establishes the relationship between the attributes interactively and categorizes the attributes into the 'Cause' group and 'Effect' group.

As per Table 20, a maximum value of (X_m+Y_n) column is 16.780 for Hazard control and safety followed by Pollution Control with a value of 16.415 and Waste handling having a value of 16.022. The degree of prominence can be arranged as Hazard control and safety (HCS) > Pollution control (PC) > Waste handling (WH) > Garage space and ambiance (GSA) > Energy and water conservation

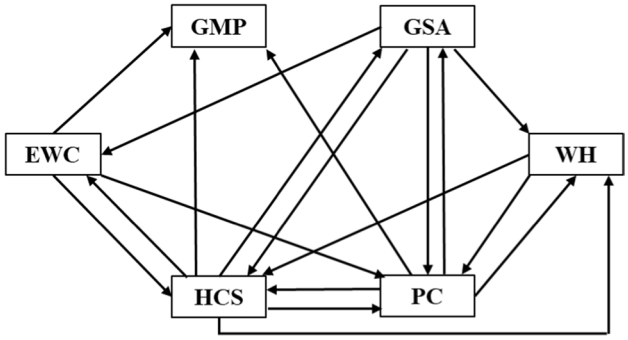


Fig. 2 Influence relationship map of SMPAs for establishing relationships

(EWC) > Garage Management Policies (GMP). The column values of $(X_m - Y_n)$, garage management policies (GMP), pollution control (PC), and waste handling (WH) indicate that they do belong to the effect group since their values are negative. However, energy and water conservation (EWC), hazard control and safety (HCS), and garage space and ambiance (GSA) have positive values for $(X_m - Y_n)$ and hence belong to the cause group. Hazard control and safety (HCS) having the highest value in $(X_m - Y_n)$ will have the greatest influence over other attributes in the effect group.

Taking the mean value of all the elements of the total relation matrix, the threshold value comes to 1.314. Now, considering the second row, i.e., garage space and ambiance (GSA), the values of pollution control (PC), waste handling (WH), energy and water conservation (EWC), and hazard control and safety (HCS) are greater than the threshold value. Hence, it can be concluded that GSA will influence pollution control (PC), waste handling (WH), energy and water conservation (EWC), and hazard control and safety (HCS). Similarly, the same relationship can be established for other rows too. These relationships among the attributes are represented by an Influence Relationship Map shown in Fig. 2.

Conclusion

Implementation of sustainable practices in manufacturing organizations is vital process for enhancing the efficiency of the organization through judicious usage of resources. The adoption of sustainable maintenance initiatives is an important issue in today's contexts as automobile maintenance activities involve a considerable amount of waste generation, energy consumption, and pollution. There various sustainable maintenance performance attributes of automobile garages were identified through literature and garage survey. These include garage management policies (GMP), garage space and ambiance (GSA), waste handling (WH), pollution control (PC), energy and water conservation (EWC), and hazard control and safety (HCS). A hybrid approach comprising AHP and PROMETHEE-II helped in ranking the garages based upon their sustainable maintenance performance and DEMATEL helped in identifying the causal relationships and relative significance among the sustainable maintenance performance attributes.

The study provided valuable information for policymakers to develop a more coherent, broader, and effective set of strategies to increase the implementation of sustainable maintenance performance attributes in automobile garages in India. Also, the results obtained will help in improved resource allocation in overcoming the challenges. The study has certain limitations as well. The structured models have their limitations such as the initial matrix is highly dependent on the judgments of the experts on the topic and there is a possibility of subjectivity in such judgments.

Acknowledgements Authors are highly grateful to the Department of Mechanical Engineering, Delhi Technological University, Delhi (India) and Department of Automobile Engineering, A.D. Patel Institute of Technology, CVM University Vallabh Vidyanagar, Gujarat (India) for their support.

Funding This research received no specific Grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Adela Y, Ambelu A, Tessema D (2012) Occupational lead exposure among automotive garage workers—a case study for Jimma town. Ethiop J Occup Med Toxicol. https://doi.org/10.1186/1745-6673-7-15

Ahmad I, Rehan M, Balkhyour M, Abbas M, Basahi J, Almeelbi T, Ismail IM (2016) Review of environmental pollution and health risks at motor vehicle repair workshops challenges and perspectives for Saudi Arabia. Int J Agric Environ Res 2(1):1–23

Ahmed S, Masjuki H, Taha Z (2005) TPM can go beyond maintenance: excerpt from a case implementation. J Qual Maint Eng 11:19–42

Ajukumar VN, Gandhi OP (2013) Evaluation of green maintenance initiatives in design and development of mechanical systems using an integrated approach. J Clean Prod 51:34–46

Amelia L, Wahab DA, Che Haron CH, Muhamad N, Azhari CH (2009) Initiating automotive component reuse in Malaysia. J Clean Prod 17:1572–1579

Apreko AA, Danku LS, Akple MS, Aboagye J (2015) Occupational health and safety management: safe work environment in the local automotive garage in Ghana. Int J Acad Res Bus Soc Sci 5:222–230

Aroonsrimorakot S (2010) Emission clinic green service: environmental management for car and motorcycle maintenance center. Environ Nat Resour 8:36–47

Ataro Z, Geremew A, Urgessa F (2018) Occupational health risk of working in garages: comparative study on blood pressure and hematological parameters between garage workers and Haramaya University community, Harar, eastern Ethiopia. Risk Manag Healthc Policy 11:35–44

Azadeh A, Gaeini Z, Moradi B (2014) Optimisation of HSE in maintenance activities by integration of continuous improvement cycle and fuzzy multivariate approach: a gas refinery. J Loss Prev Process Ind 32:415–427

Bhatia MS, Srivastava RK (2018) Analysis of external barriers to remanufacturing using grey-DEMATEL approach: an Indian perspective. Resour Conserv Recycl 136:79–87

Bouguerra S, Chelbi A, Rezg N (2012) A decision model for adopting an extended warranty under different maintenance policies. Int J Prod Econ 135:840–849

Bourne M, Neely A, Mills J, Platts K (2003) Implementing performance measurement systems: a literature review. Int J Bus Perform Manag 5:1–24

Brans JP, Vincke P (1985) Note-a preference ranking organization method: the PROMETHEE method for multiple criteria decision-making. Manag Sci 31:647–656

Brans JP, Vincke P, Mareschal B (1986) How to select and how to rank projects: the PROMETHEE method. Eur J Oper Res 24:228–238

Brans JP (1982) L'ingénièrie de la decision; elaboration d'instrumentsd'aide à la décision. La méthode PROMETHEE. Presses de l'Université Laval, pp 183–213

Brito EPZ, Aguilar RLB, Brito LAL (2007) Customer choice of a car maintenance service provider: a model to identify the service attributes that determine choice. Int J Oper Prod 27:464–481

Chaisrisuk H, Charoensaeng A, Khaodhiar S (2019) Hazardous waste management from garages in Bangkok by material flow analysis (MFA). J Ind Intell Inf 7:33–36

Chan YCL (2006) An analytic hierarchy framework for evaluating balanced scorecards of healthcare organizations. Can J Adm Sci 23:85–104

Chiu MC, Hsieh MC (2016) Latent human error analysis and efficient improvement strategies by fuzzy TOPSIS in aviation maintenance tasks. Appl Ergon 54:136–147

Coetzee JL (1998) Reliability-centered maintenance. Maintenance Publishers, Hatfield, Republic of South Africa

Cruickshank B (2010) Occupational Health Management. Warwick Medical School, UK

Dant T, Bowles D (2003) Dealing with dirt: servicing and repairing cars. Sociol Res Online 8:1–31

Dekker R (1996) Applications of maintenance optimization models: a review and analysis. Reliab Eng Syst Saf 51:229–240

Demie G (2015) Analyzing soil contamination status in garage and auto mechanical workshops of Shashemane city: implication for hazardous waste management. Environ Syst Res 4:15

Dhillon BS (2014) Human error in maintenance: an investigative study for the factories of the future. In: 27th International Conference on CAD/CAM, Robotics and Factories of the Future, July 22–24, Middlesex, London, UK, pp 1–13

Doshi JA, Panchal DU, Maniar JP (2014) Vehicle maintenance and garage practice. PHI Learning Private Limited, New Delhi, India

Edgar E (2009) An unusual obituary: the integrated waste management board. Capitol Weekly (Sacramento, CA), US

El-Saadawy MS, Attwa EM, El-Tayeb IM, Zalat MM (2011) Dermatoses and hematological disorders among car mechanics in Zagazig city and their effects on quality of life. Zagazig Univ Med J 17:142–156

Henderson K, Pahlenkemper G, Kraska O (2014) Integrated asset management: an investment in sustainability. Proc Eng 83:448–454

Hennequin S, Restrepo LMR (2016) Fuzzy model of a joint maintenance and production control under sustainability constraints. IFAC Pap Online 49:1216–1221

Hunt J, Cathy C, Michael P, Anthony M (2000) Occupation-related burn injuries. J Burn Care Rehab 21:358–390

Ighravwe DE, Oke S (2017a) A multi-hierarchical framework for ranking maintenance sustainability strategies using PROMETHEE and fuzzy entropy methods. J Build Pathol Rehab 21:9. https://doi.org/10.1007/s41024-017-0028-7

Ighravwe DE, Oke SA (2017b) Ranking maintenance strategies for sustainable maintenance plan in manufacturing systems using fuzzy axiomatic design principle and fuzzy-TOPSIS. J Manuf Technol Manag 28:961–992

Ijeoma K, Achi OK (2011) Industrial effluents and their impact on water quality of receiving rivers in Nigeria. J Appl Technol Environ Sanit 1:75–86

Ikwan FC, Sanders D, Haddad M (2020) A combined AHP-PROMETHEE approach for intelligent risk prediction of leak in a storage tank. Int J Reliab Qual Saf Eng. http://www.ijrrs.com/article_119488.html

ILO (2000) International Data Sheet on Occupation. Automobile Mechanics, ILO, Geneva

Ishola AB, Okechukwu IM, Ashimedua UG (2017) Serum level of lead, zinc, cadmium, copper and chromium among occupationally exposed automotive workers in Benin City. Int J Environ Pollut Res 5:70–79

ISO-14040 (2006) Environment management-life cycle assessment-e principles and framework. www.iso.org

James AT, James J (2020) Service quality evaluation of automobile garages using a structural approach. Int J Qual Reliab Manag. https://doi.org/10.1108/IJQRM-03-2020-0066

James AT, Gandhi OP, Deshmukh SG (2017) Development of methodology for the disassemblability index of automobile systems using a structural approach. Proc Inst Mech Eng D J Automob Eng 231:516–535

James AT, Kumar G, Bhalla M, Amar M, Jain P (2020) Analysis of challenges for automobile service garages in India: a structural modeling approach. J Adv Manag Res. https://doi.org/10.1108/JAMR-04-2020-0059

Kant I, Notermans J, Borm P (1990) Observations of working postures in garages using the Ovako working posture analyzing system (OVVAS) and consequent workload reduction recommendations. Ergonomics 33:209–220

Kelly A (1997) Maintenance strategy. Butterworth-Heinemann, Oxford, UK

Kim HJ, Raichur V, Skerlos SJ (2008) Economic and environmental assessment of automotive remanufacturing alternator case study. In: Proceedings of the 2008 International Manufacturing Science and Engineering Conference, October 7–10, Illinois, USA, pp 33–40

Kumar A, Dixit G (2018) An analysis of barriers affecting the implementation of e-waste management practices in India: a novel ISM-DEMATEL approach. Sustain Prod Consum 14:36–52

Kumar U, Galar D, Parida A, Stenstrom C (2013) Maintenance performance metrics: a state-of-the-art review. J Qual Maint Eng 19:233–277

Lind S, Neonen S (2008) Occupational risks in industrial maintenance. J Qual Maint Eng 14:194–204

Liu M (2005) Multi objective maintenance planning optimization for deteriorating bridges considering condition, safety, and life-cycle cost. J Struct Eng 131:131–833

Mansur AV, Brondízio ES, Roy S, Hetrick S, Vogt ND, Newton A (2016) An assessment of urban vulnerability in the Amazon Delta and Estuary: a multi-criterion index of flood exposure, socio-economic conditions and infrastructure. Sustain Sci 11:625–643

Martorell S, Sanchez A, Carlos S, Serradell V (2002) Simultaneous and multi-criteria optimization of TS requirements and maintenance at NPPs. Ann Nucl Energy 29:147–168

Mathiyazhagan K, Sengupta S, Poovazhagan L (2018) A decision-making trial and evaluation laboratory approach to analyse the challenges to environmentally sustainable manufacturing in Indian automobile industry. Sustain Prod Consum 16:58–67

Motey F, Coast C (2013) Analysis of safety practices by automobile mechanics at Siwdo Kokompe. Ind Eng Lett 3:1–18

Muchiri P, Pintelon L, Gelders L, Martin H (2011) Development of maintenance function performance measurement framework and indicators. Int J Prod Econ 131:295–302

Muttamara S, Alwis K (1994) Health impacts of garage workers: a preliminary study. J Environ Health 56:19–24

Occupational Safety and Health Council (2011) Guideline for good occupational hygiene practice in workplace. Available at: http://www.labour.gov.hk. Accessed 22 Mar 2020.

Parker DL, Bejan A, Brosseau LM (2012) A qualitative evaluation of owner and worker health and safety beliefs in small auto collision repair shops. Am J Ind Med 55:474–482

Pathberiya R, Dasanayaka SWSB, Serhan OA, Roudaina H (2017) Environmental and health impact of small garages and workshops: a case study based on Sri Lanka. Prog Ind Ecol 11:314–342

Pintelon LM, Van Wassenhove LN (1990) A maintenance management tool. Omega 18:59–70

Quan G, Greenwood GW, Liu D, Hu S (2007) Searching for multi objective preventive maintenance schedules: combining preferences with evolutionary algorithms. Eur J Oper Res 177:1969–1984

Ramanathan R (2001) A note on the use of the analytic hierarchy process for environmental impact assessment. J Environ Manage 63:27–35

Reason J (1997) Managing the risks of organizational accidents. Ashgate Publishing Limited, London, UK

Reason J, Hobbs A (2003) Managing maintenance error—a practical guide. Ashgate Publishing Company, Burlington, USA

Resat HG, Unsal B (2019) A novel multi-objective optimization approach for sustainable supply chain: a case study in packaging industry. Sustain Prod Consum 20:29–39

Rongo L, Barten F, Msamanga G, Heederik D, Dolmans W (2004) Occupational exposure and health problems in small-scale industry workers in Dares Salaam, Tanzania: a situation analysis. Occup Med 54:42–60

Saaty TL (1994) Fundamentals of decision-making and priority theory with AHP. RWS Publications, Pittsburg, California, USA

Saaty TL (2008) Decision making with the analytic hierarchy process. Int J Serv Sci 1:83–98

Sabharwal N, Soch H, Kaur H (2010) Are we satisfied with incompetent services? A scale development approach for service recovery. J Serv Res 10:125–142

Safar A, Abbaspour M, Javid AH (2020) The application of multi-criteria (AHP-PROMETHEE) decision-making methods in selecting and prioritizing the green area irrigation resources. Int J Environ Sci Technol. https://doi.org/10.1007/s13762-020-02875-x

Sangodoyin AY (1991) Groundwater and surface pollution by open refuse dump in Ibadan. Niger J Discov Innov 3:24–31

Sari E, Shaharoun AM, Ma'aram A, Yazid AM (2015) Sustainable maintenance performance measures: a pilot survey in Malaysian automotive companies. Proc CIRP 26:443–448

Schwartz E (1987) Proportionate mortality ratio analysis of automobile mechanics and gasoline service station workers in New Hampshire. Am J Ind Med 12:91–99

Seitz MA (2007) A critical assessment of motives for product recovery: the case of engine remanufacturing. J Clean Prod 15:1147–1157

Sheng-Li S, Xiao-Yue Y, Hu-Chen L, Zhang P (2018) DEMATEL technique: a systematic review of the state-of-the-art literature on methodologies and applications. Math Probl Eng 2018:1–34

Sichinga MC, Frazee J, Tong AZ (2016) Efficiency and kinetics in treatment of wastewater from garages and residential oil spills using membrane bioreactor technology. Int J Environ Sci Technol 13:135–146

Simoes JM, Gomes CF, Yasin MM (2011) A literature review of maintenance performance measurement: a conceptual framework and directions for future research. J Qual Maint Eng 17:116–137

Swanson L (2001) Linking maintenance strategies to performance. Int J Prod Econ 70:237–244. https://doi.org/10.1016/S0925-5273(00)00067-0

Tian H, Wang ZH (2020) Chinese green process innovation in automotive painting: the strategic niche management perspective. Int J Environ Sci Technol 17:993–1010

Tuzkaya G, Ozgen A, Ozgen D, Tuzkaya UR (2009) Environmental performance evaluation of suppliers: a hybrid fuzzy multi-criteria decision approach. Int J Environ Sci Technol 6:477–490

Tzeng G, Huang J (2011) Multiple attribute decision making methods and applications. CRC Press, Taylor and Francis Group, A Chapman & Hall Book, Boca Raton

Van Horenbeek A, Pintelon L (2014) Development of a maintenance performance measurement framework—using the analytic

network process (ANP) for maintenance performance indicator selection. Omega 42:33–46
Vujanovic D, Momcilovic V, Bojovic N, Papic V (2012) Evaluation of vehicle fleet maintenance management indicators by application of DEMATEL and ANP. Expert Syst Appl 39:10552–10563
Vyas H, Das S, Mehta S (2011) Occupational injuries in automobile repair workers. Ind Health 49:642–651
Wang B, Levrat B (2014) Advanced maintenance services for promoting sustainability. Proc CIRP 22:15–22
Wireman T (2005) Developing performance indicators for managing maintenance, 2nd edn. Industrial Press, New York, USA
Yasin S, Iqbal T, Arshad Z, Rustam M, Zafar M (2012) Environmental pollution from automobile vehicle service stations. J Qual Technol Manag 8(1):61–70

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

Alexandria Engineering Journal (2022) xxx, xxx–xxx

HOSTED BY

ELSEVIER

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com



ORIGINAL ARTICLE

Augmentation in performance and security of WSNs for IoT applications using feature selection and classification techniques

Rajiv Yadav [a,*], **Indu Sreedevi** [a], **Daya Gupta** [b]

[a] *Dept of Electronics and Communication Engineering, Delhi Technological University, Delhi 110042, India*
[b] *Dept of CSE Engineering, Delhi Technological University, Delhi 110042, India*

Received 12 July 2022; revised 13 September 2022; accepted 10 October 2022

KEYWORDS

WSNs;
Feature selection;
Classification techniques;
NSL-KDD dataset;
Accuracy;
IoT

Abstract Energy consumption, secure connection of Sensor Nodes (SNs), and their performance analysis play a major role in Wireless Sensor Networks (WSNs) research. Increased case use of various Internet of Things (IoT) applications has led to the development of complicated networks. To address the threats and security issues faced in complicated WSNs, a novel combined feature selection method known as the Fast Correlation-based Feature Selection (FCBFS) with XG-Boost has been proposed for the NSL-KDD intrusion detection benchmark dataset. It helps in the selection of the best features in a cluster-based WSN before the classification step. Five popular machine learning-based classifiers, namely decision tree, random forest, Naïve Bayes, extra tree, and XG-Boost, are used for the development of a robust intrusion detection system in WSN and its IoT applications. The effectiveness and robustness of the developed ensemble methods have been checked using classification accuracy, precision, recall, and F-score. XG-Boost classifier along with the FCBFS process has performed best with a classic accuracy, precision, recall, and F-score up to 99.84%, 99.83%, 99.84%, and 99.82% on multiple runs. Upon comparison with the existing state-of-the-art work in this field, the proposed work has outperformed. Results show that, in contrast to the previous filter approaches, our proposed process can successfully minimize the number of features while maintaining a high classification precision and recognition rate. It further tends to lower the overall energy required by sensor nodes during attack detection, hence extending the network lifetime and usefulness to a sufficient time frame.

© 2022 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Alexandria University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

WSN is an assembly of robust sensor networks that exchange data through a wireless system. A sensor is a device that collects information about a physical thing or an incident. The sensor center is made up of recognition, actuation, and power

* Corresponding author.
E-mail addresses: yadav.rajiv217@gmail.com (R. Yadav), s.indu@dce.ac.in (I. Sreedevi), dgupta@dec.ac.in (D. Gupta).

Peer review under responsibility of Faculty of Engineering, Alexandria University.

https://doi.org/10.1016/j.aej.2022.10.033

1110-0168 © 2022 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Alexandria University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

elements that are all synchronized on one or more sheets that make up the embedded system. A sensor network employs several sensors to communicate data to a designated location over the internet [1]. The SN's sensor element detects the observable status of a scenario or physical environment, while the microprocessor guarantees that the data is effectively calculated. The node's wireless radio also takes guarantees communication with nearby nodes. Sensors are frequently installed in distant, rugged, and unsupervised locations for extended periods. Node preservation after the deployment of sensors is difficult due to the inaccessibility of these locations. Temperature, sound, vibration, and weight are all detectable by them. The application of WSNs is becoming widespread with the growing use of the IoT, supervision, and distant monitoring of mobile phones [2,3].

WSNs are frequently employed for a variety of purposes, such as covering military applications, environmental catastrophes, animal monitoring, safety inspections, etc. [4]. The basic goal is to assemble useful statistics in a steady situation. Because there are multiple barriers to the security of these networks and IoT, providing and sustaining data protection is a critical concern. Distributed Denial of Service (DDoS) assaults, server negotiation caricature attacks, and protocol-specific attacks are all examples of these types of attacks. A DDoS assault can swiftly deplete available resources through a variety of methods, including flooding [5]. The risk is far higher when it comes to military and medical uses. Presenting a security model that considers these limits while providing security is a key task nowadays. Extortion has been recognized as one of the key reasons for such assaults, which have continued to grow in magnitude and sophistication.

Various Intrusion Detection Systems (IDSs) have been suggested by scholars to safeguard the network from adversary intrusion [6]. The first uses signatures of known attack patterns, whereas the latter uses data mining, Machine Learning (ML), and statistical modeling to profile a statistical use model. Signature-based detection has a noteworthy weakness that cannot detect unknown assaults. As a result, a hybrid approach has emerged, which combines the complementing features of both methodologies to attain a greater detection rate.

Feature Selection (FS) helps to improve simplification by decreasing over-fitting, computational efficiency, and stuffing, as well as improving cataloging precision by developing an accurate prediction model, as shown in Fig. 1.

The FS procedure is divided into four steps [7], as discussed below:

- Feature subset generation: It is an exploration technique that uses one of the succeeding search strategies to pick a feature subset: full search, heuristic search, and random search.
- Feature evaluation: Filters, wrappers, and embedded designs are utilized to analyze the created subsets in this item.
- Stopping criterion: When the optimization process should be stopped, this feature determines. The determined number of repetitions or defining a goal objective, such as absolute classification accuracy, are examples of stopping criteria.
- Validation: When employing the created subsets in an optimization method, this procedure verifies the correctness of the results.

The rapid increase in advanced technologies, like mobile communications, device availability, and processing systems, has led to the next era of interaction known as the Internet of Things. As a result, IoT security is a serious concern in safeguarding the hardware and networks of IoT systems [8]. Security has not been considered in the creation of networking gadgets because they are still relatively recent inventions. Academic research on confidentiality problems for IoT devices has shown encouraging breakthroughs in recent years. The tactics and security processes now available are primarily based on classic network security approaches. Establishing security mechanisms in an IoT system is more complex than in a traditional network due to the wide variety of technologies and devices, as well as the scalability or number of nodes in the network.

WSN has vast real-world applications for IoT [8] in the following areas:

- Defense: In the battlefield for surveillance and monitoring.
- Environment: Forest fire detection, Pollution monitoring, and early detection of earthquakes.
- Health Care: For patient diagnosis and monitoring.
- Agriculture: Crop monitoring, soil monitoring, temperature, and humidity.
- Internet of Things (IoT): Home automation, Smart cities, and smart grids.

In this paper, the authors introduce an FS phase using an FCBFS method to select almost one-fourth of the original dataset (10 most important features) before classifying it with different classifier algorithms. Because SNs have a limited number of resources, feature selection is used in ML

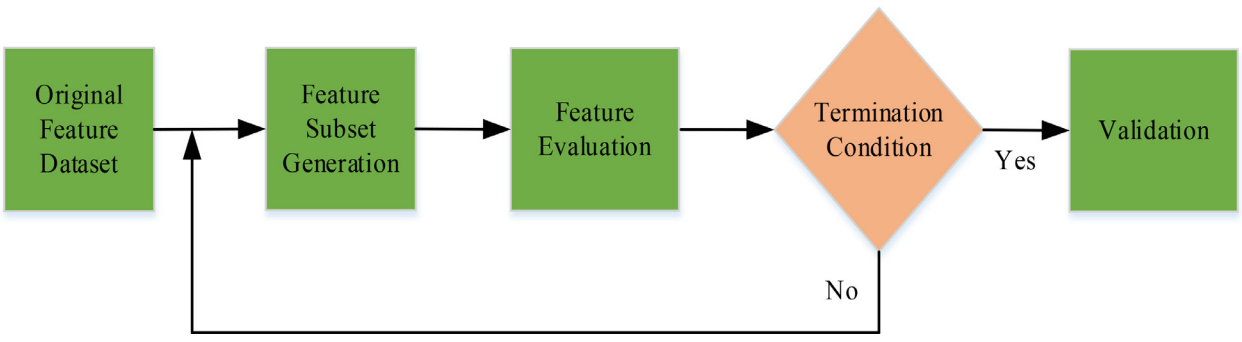


Fig. 1 Feature selection process.

approaches to minimize duplicate characteristics and lower the difficulty of the projected scheme. The IDS benchmark dataset, NSL-KDD, with 41 characteristics, was used to test the performance taking into account precision, classification accuracy, recall, and F-score [9]. In comparison to [10], the findings reveal that our suggested strategy successfully reduces the number of characteristics while retaining excellent classification accuracy and precision, as well as F-score.

In feature engineering, there are various challenges in training ML models. Updating in or deleting incomplete information, encrypting category parameters, changing numeric attributes, and extracting information from dates, GPS locations, text, and photos are just a few of the many tasks involved. It takes longer to examine and prepare the data we are working with; hence it is time-consuming. It also includes computation complexity, repetitiveness, reproducibility, and interpretability. Our suggested method is designed to shorten the time required for data modification by avoiding writing repeated code while still guaranteeing repeatability and delivering features that are understandable.

By proposing a lightweight method, the IDS's complexity is reduced. Because of the compact IDS involvedness, the SNs will use a smaller amount of energy while retaining good availability. Because the SNs are battery-powered, it's critical to keep the network up and running for as long as possible. As a result, our suggested IDS protection method is appropriate for usage in a real-time WSN since it effectively extends network lifetime and usefulness [11].

This research proposes an FCBFS method for extracting features and enhancing detection accuracy. This research article's contribution is concise as follows:

- The status of an FS technique is deliberated and experimented with an ensemble method.
- A novel filtering strategy is presented for the FS process in the proposed IDS's.
- Effective comparison of different machine learning algorithms, namely decision tree, random forest, Naïve Bayes, extra tree, and XG-Boost, is carried out. The best-achieved metrics are equated with the current state-of-the-art work in this field.

2. Related work

The researchers in [12,13] presented a faster DNN architecture for detecting anomalies in network data. The NSL-KDD dataset is used to calculate the training duration and evaluate the detection mechanism's efficacy. These studies focus on using an accelerator platform to handle large datasets and detecting complicated relationships in the input dataset to identify distinct attack types. One of the major goals of these researches is to compare the act of DNN training on diverse machine types and core counts. The authors [14,15] proposed a customized dataset for WSN to aid in the exposure and organization of four types of DoS attacks: Blackhole, Gray-hole, Flooding, and Scheduling assaults. This research looks at the LEACH protocol, which is one of the most extensively used cluster-based routing methods in WSNs. Using technology, data from Network Simulator 2 (NS-2) was gathered and processed to give 23 characteristics. This study emphasizes the need to address security early in the network protocol development process. Without this, hostile attacks would progressively target inherited weaknesses in various network protocols and other software.

The authors in [16,17] proposed a Restricted Boltzmann Machine (RBM) and a deep belief network to build a deep learning strategy for anomaly identification. To conduct unsupervised feature reduction, the authors employ a one-hidden layer RBM. This RBM's weights are transmitted to another RBM, resulting in a deep belief network. A fine-tuning layer that contains a Logistic Regression (LR) classifier with multi-class soft-max feeds the pre-trained weights. They developed a multi-class LR layer and trained using 10 epochs on the enhanced pre-trained data to improve the network's overall performance. They also created a network with a fast-training period and little dataset pre-processing. The authors in [18,19] verified a range of machine learning methods, including neural networks, that leverage IoT-specific network characteristics to influence FS that may result in high accuracy DDoS detection in IoT networks. The authors demonstrated in this paper that ML at the packet level is possible. DoS detection from consumer IoT devices can accurately discriminate between regular and DoS attack traffic. They also used a limited feature set to lower the computational cost, which is crucial for real-time classification and middlebox deployment.

In [20,21] Based on the C.4.5 algorithm, the authors developed a DDoS detection system to combat the attacks. When paired with signature detection techniques, this strategy creates a decision tree that can automatically and effectively identify signature assaults for DDoS flooding attacks. A comparison of several machine learning algorithms and methodologies was offered. The authors concluded from the data that our suggested solution, which uses the C4.5 algorithm to perceive DDoS assaults, offers more precise outcomes than existing ML algorithms. The authors in [22,23] suggested a layered hierarchical approach for identifying abnormalities in various online social networks. The experiment was conducted using four data sets, with Twitter and Facebook showing the best outcomes in anomaly detection. Existing solutions either focus on a fixed collection of named data or monitor the behavior of randomly selected hubs rather than the unstructured flow of data in interpersonal groups. Additionally, the characteristics employed in the mining process might be expanded to provide more precise findings.

The researchers in [24,25] created a clustering-based approach for separating data describing network traffic patterns, including both normal and DDoS activities. The characteristics are used to identify assaults at the victim's end, and the process is shown with three features that may be observed in the target system. To label the data and establish classes to separate assaults from normal traffic, a voting approach is also offered. Using a demonstrative fraction of the benchmark CICIDS2017 dataset and additional vectors of attack, the technique is also verified for label assignment accuracy. The authors in [26,27] used the ping of death approach to launch a DDoS assault, which they discovered using an ML technique using the WEKA tool. This experiment uses the NSL-KDD dataset. The categorization of the normal and assault samples was done using a random forest technique. The model was trained using a random forest approach.

Authors in [28] suggested a partial channel connection for Differentiable Neural Architecture Search (ADARTS) based

Please cite this article in press as: R. Yadav et al., Augmentation in performance and security of WSNs for IoT applications using feature selection and classification techniques, Alexandria Eng. J. (2022), https://doi.org/10.1016/j.aej.2022.10.033

on channel attention. In this method, selected routes are picked through the learning algorithm and delivered into the processing region, while the remaining channels are instantly associated with the treated channels. The efficiency of the search process and the use of memory may both be significantly increased by choosing a small number of channels with greater attention weights. When there are too many skip connections during the selection phase, ADARTS can adequately address this issue and provide network architectures that function better.

Authors in [29,30] proposed a Non-dominated Sorting Genetic Algorithm-III (NSGA-III) as a potential solution to the problem of missing data in practical applications. The mean imputation approach was utilized to cope with the missing data by the authors, who also identified dependability as a concern. The suggested model, when linked with NSGA-III, effectively handles the FS problem, according to experimental findings. A Self-adaptive Particle Swarm Optimization (SaPSO) approach has been suggested specifically for large-scale feature selection. The SaPSO first employs an encoding approach for the feature selection issue. The investigation of three crucial self-adaptive algorithm-related difficulties follows. The algorithm is then suggested, which features a standard self-adaptive technique. The experimental findings on 12 datasets demonstrate that the proposed algorithm's solution size is always less than that of its computing equivalents.

Authors in [31,32] proposed a route selection strategy based on the protection of network connections developed in this research to increase the total network lifespan. Both proactive and reactive route maintenance procedures have been used to mimic the proposed Variable Data-rate Routing (VDR) approach to route selection. The simulation results demonstrate that the proposed method outperforms the current State-of-the-Art energy-efficient route selection algorithms in terms of network lifetime, energy consumption, total network connectivity, and packet delivery ratio. For the multiple route discoveries, the authors also presented an Ad hoc On-demand Multipath Distance Vector (AOMDV) protocol together with K-Nearest Neighbor (KNN). The Advanced Encryption Standard (FAES) encryption algorithm is the one that is proposed to be utilized for node selection and cryptography.

Authors in [33,34] proposed Multi-objective Ant colony optimization (MOACO) with the Q-learning method to improve inter-cluster data aggregation. The majority of IoT network models feature data acquisition, energy-constrained devices, and utilization-enhancing optimization strategies. The simulation studies reveal that, when compared to the ABC and GA, the proposed method analyses better throughput, longevity, and energy enhancement. The AI and MOACO algorithms are also used to assess the jamming activity at low, high, and moderate levels. Studies may be improved to deal with the energy and security factors. The suggested model may be studied with the energy parameters utilizing conventional and cutting-edge machine learning approaches, which may add additional complexity to execution in the jamming assault.

3. Methodology

3.1. Dataset description and pre-processing

The collection of significant features from the dataset is critical to the performance of any ML model. The most common NSL-KDD dataset from [35] is utilized for classification in the implementation. There are 41 distinct attributes in this dataset, including content type, basic kind, and traffic type, as shown in Table 1. Each record reveals diverse flow aspects, each of which is categorized as an attack type or normal.

The 42nd attribute offers evidence of the five unlike types of system connection vectors. They are further separated into one regular class and four attack classes. DoS, Probe, R2L, and U2R are the four attack categories that are divided into four assault types [37]. The attack classes are described below:

- Denial-of-Service attack (DoS): The attacker attempts to stop authorized users from utilizing a service. The earliest type of cyber squeezing assault is a DoS attack in which the invader makes the server extremely busy, and as a result of that program, the legitimate user is denied access. This kind of attack frequently uses the flood attack strategy, which is making unauthorized system and network requests to get the targeted person's fair service denied by the computer system or network. The attack types include apache, back, land, mailbomb, neptune, pod, processtable, smurf, teardrop, and udpstorm.
- Probe: The attacker attempts to learn more about the intended host. This attack is an effort to fold data about a system of PCs with the clear goal of evading its security protections. Attacker-supplied scanning traffic can check the computer for vulnerabilities when it links to the corresponding targeted device, making it simpler to select appropriate tools and targets for further attacks. The attack types include psweep, mscan, nmap, portsweep, saint, and satan.
- Remote to local attack (R2L): The attacker tries to obtain access since they do not have an account on the victim's computer. This attack entails sending particular information to the target across the network, then obtaining approval to access the victim's device to gain access to the victim's terminal structure for further assault. The FTP client then mistakenly executes one or more commands

Table 1 Dataset description [36].

Criteria	Description
Dataset	NSL-KDD
Number of records	1,49,470
Number of attack categories	4 (DoS, Probe, R2L, U2R)
Number of network features	41
Type of network features	Basic features, traffic features, host features, and content features.
Number of records in the training set	1,26,620
Number of records in the test set	22, 850

included in the directory's excessively long name with the user's consent. The attacker encroaches into a distant appliance and acquires local access to the victim mechanism through an illegal entree from a remote appliance. The attack types include ftp_write, guess_password, imap, and multihop.

- Unauthorized access to local super-user (root) privileges (U2R): The attacker attempts to get super-user capabilities while having local access to the victim's computer. A regular account is utilized by a U2R attacker to log into the target system. The attacker attempts to obtain root/administrator rights by manipulating the softness of the victim's regular account. Once the attacker has gained access to the victim's terminal structure, they will exploit the attack circulation to change normal users' permissions to system access rights until finally obtaining administrator or root credentials. It gains access to the system's roots and launches a series of assaults and illegal efforts i.e., buffer_overflow, loadmodule, perl, rootkit, ps, sqlattack, and xterm.

The most important phase in the experiment is the model step because the goal of this study is to treasure the best FS technique and ML model for dealing with the challenges, as shown in Fig. 2.

Data pre-processing is the most significant phase in every ML project. The basic goal of data cleaning is to assure data accuracy by searching for duplicated or unrealistic information. However, there are numerous non-numeric properties of data collection. In general, Scikit-Learn classifiers can handle numeric characteristics well but not non-numeric attributes. Rather than merely converting non-numeric qualities to numeric attributes, it is important to use a suitably specialized encoding mechanism to pre-process the data in the data processing. The computer cannot match one attribute to another since they are all discontinuous and unordered, but it may place N potential outcomes of a conceivable attribute in N registers using binary to store the data after applying a certain encoding method. Each state has its own set of register bits, with only one active at any one moment. In this experiment, one-hot coding is used. In pre-processing, the raw dataset is cleaned to convert it into a set-up that can be utilized to train an ML model. When creating a well-organized model, the data must be in a precise format.

3.2. Feature selection

The goal of FS is to eliminate material that is redundant or useless. Not every data aspect makes sense. This is a method for choosing a subset of important attributes that accurately describes a situation. In addition, the selection will reduce the amount of data picked. It's critical to choose the right characteristics from the complete dataset, as shown in Table 2, because they can impact the model's performance and accuracy.

To begin with, unconnected traits may signal some level of effectiveness to a certain situation, but this does not imply that they are inextricably linked. As a result, suitable models cannot be built appropriately. This will not only damage the data's correctness and model's dependability, but it will also cause overfitting issues. Second, the bigger the measurement of the source information, the lengthier the dispensation and computing time for the classifier will be. The innovative data includes a large number of dimensions, which increases

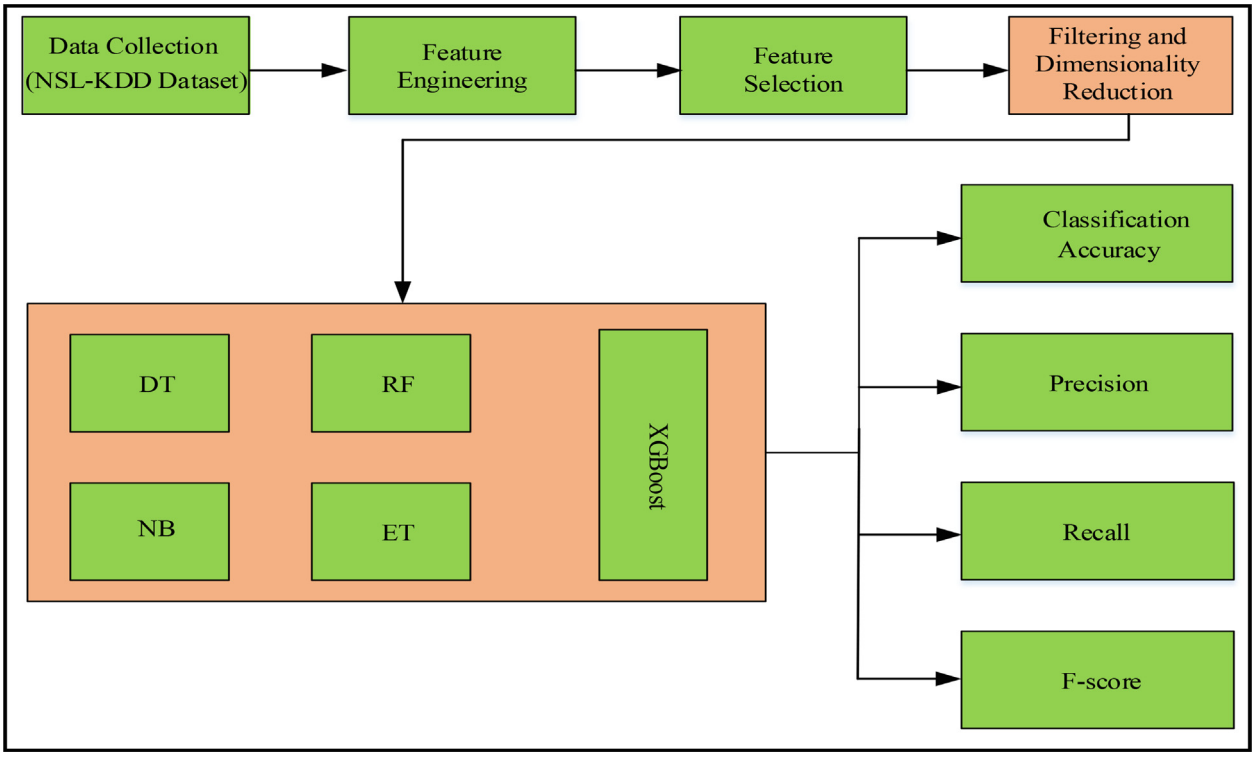


Fig. 2 Experimental flow of proposed work.

Table 2 NSL-KDD dataset features [38].

No.	Dataset features	No.	Dataset features	No.	Dataset features
1.	Duration	15.	Su_attempted	29.	Same_srv_rate
2.	Protocol_type	16.	Num_root	30.	Diff_srv_rate
3.	Service	17.	Num_file_creations	31.	Srv_diff_host_rate
4.	Flag	18.	Num_shells	32.	Dst_host_count
5.	Src_bytes	19.	Num_access_files	33.	Dst_host_srv_count
6.	Dst_bytes	20.	Num_outbound_cmds	34.	Dst_host_same_srv_rate
7.	Land	21.	Is_host_login	35.	Dst_host_diff_srv_rate
8.	Wrong_fragment	22.	Is_guest_login	36.	Dst_host_same_src_port_rate
9.	Urgent	23.	Count	37.	Dst_host_srv_diff_host_rate
10	Hot	24.	Srv_count	38.	Dst_host_serror_rate
11.	Num_failed_logins	25.	Serror_rate	39.	Dst_host_srv_serror_rate
12.	Logged_in	26.	Srv_serror_rate	40.	Dst_host_rerror_rate
13.	Num_compromised	27.	Rerror_rate	41.	Dst_host_srv_rerror_rate
14.	Root_shell	28.	Srv_rerror_rate		

machine learning time, lowers efficiency, and diminishes prediction efficiency. The impact of time on intrusion detection difficulties is unbearable.

Reducing duplication and picking more appropriate characteristics aids in:

- Reduces the dataset's size.
- Reduces the likelihood of data being manipulated.
- Reduces the chance of overfitting.
- Reduces the amount of time it takes to train the model.
- Enhances the model's accuracy and reliability.

Pseudocode of the proposed technique (using Fast Correlation-based Feature Selection (FCBFS):

Begin: Data collection;

Input: Set of 41 features from the NSL-KDD dataset.
Output: Best Selected Features Subset.
: Select the features having k = 10; and apply Fast Correlation-based Feature Selection.
2: Calculate Pearson's Correlation Coefficient using equation (2).
3: Select the features subset which satisfies the threshold.
4: Repeatedly apply the feature selection (FCBFS) on the features obtained from step3.
5: Train and test the model.
6: End.

The most crucial stage is to remove unnecessary, redundant characteristics from the dataset. Studies suggest that the development of ML models using irrelevant characteristics or feature affect the evaluation metrics of the model. In this paper, we will use the FCBFS method that automatically chooses the ten most relevant features from the 41 features of the NSL-KDD dataset. The concept of predominant correlation is used in the algorithm. High correlation with the target and low correlation with other variables are chosen as the attributes. It substitutes a statistic known as symmetrical uncertainty (SU) instead of the conventional Pearson or Spearman correlations. The Shannon entropy and information gain principles serve as the foundation for the SU, which is based on information theory. Above a certain SU threshold, the algorithm chooses characteristics that are associated with the target. Then it finds the traits that most frequently correlate with the target. We begin the iteration with feature X, which is the highest-rated, once features are ranked. The characteristics that have a higher correlation with X than the target is eliminated. Any characteristics that have a lower correlation with X than the goal is preserved. The second-best feature on the trimmed list is the feature that the algorithm moves on to next.

Information Gain (IG) is one of the filter FS strategies used in selecting appropriate qualities from a set of features. When the value of a feature is unknown, IG reduces the uncertainty associated with determining the class attribute. It is based on information theory, and it is used to rank and choose top features to minimize feature size before the learning process begins. The information gain value of the dataset's features is rated. A characteristic with a low information gain value indicates that it has little impact on data categorization. As a result, features with small information gain values can be discarded without having a significant impact on the classification model's performance. Eq. (1) may be used to compute the information gain value for class C and input feature X.

$$IG(C, X) = H(C) - H\left(\frac{C}{X}\right) \tag{1}$$

Here, H(C|X) is the probability of class C for a given feature X, while H(C) is the entropy of class C.

3.2.1. Fast Correlation-based feature selection (FCBFS):

The dataset is first subjected to the FCBFS approach. Pearson's Correlation Co-efficient (PCC) is used to create a correlation matrix. The PCC is derived by dividing the covariance of two values and then multiplying the divided value by the standard deviation of each value, as shown in Eq. (2).

$$Coefficient = \frac{covariance(x, y)}{(stdv(x) * stdv(y))} \tag{2}$$

Here, x represents data and y is a random variable.

The link between the attributes may be understood using these coefficients. The coefficients' values are always between

−1 and 1. A symmetric matrix is formed by evaluating this connection and placing those standards in that matrix. Correlating characteristics are discarded as a consequence of this observation. Before eliminating the linked features, there were 41 features; after lowering them, only ten features remained.

3.3. Feature scaling

As the feature might comprise big values, the model must be trained on larger standards, which can have a greater impact on the outcome; feature scaling is a necessary step in creating an effective ML model. Pre-processing just prepared the data for input into the model, whether the model's conclusions are valid. Whether the subsequent states can be predicted properly is dependent on the data scaled [37]. Each piece of data contains numerous properties, yet when transformed into a number, each feature has a different assortment of standards. The normalizing technique is used to scale features. It is a method of transforming data points into the 0–1 range while maintaining their relationship. Data with a greater worth has a sophisticated percentage, which has a direct impact on data training accuracy. As a result, you must average each of these values, then subtract the mean from the eigenvalue before dividing by their standard deviation. This phase is critical for improved accuracy, and it must be completed before the model is trained.

3.4. Classification techniques and execution process

The classification phase is the last step in the development of the intrusion system. In this step, a labeled training dataset is used for experimentation and training the classifier with known labels. Later the effectiveness and robustness of the classifier are checked upon the testing dataset, which doesn't contain obvious labels. The training and testing stages of the model that employ classification-based algorithms can be separated into two categories. Labeled data are used in the training step to acquire a specific classifier. This classifier may then be used in the test step to determine if a test instance is normal or anomalous. Some classifiers like DT, RF, NB, ET, and XGBoost are used in our experiment with the NSL-KDD dataset [36 39].

3.4.1. Decision tree (DT) Classifier:

To forecast the value of the target variable, a DT classifier constructs a model that acquires basic decision rules using data attributes. A source node and many leaf nodes make up a DT. The leaf node relates to decision outcomes, whereas other nodes agree to feature tests. The sample set is partitioned into child nodes based on the outcome of the attribute test. The whole set of examples may be found at the root node. Each node's route from its root to its leaf resembles a series of choice tests. The goal of DT learning is to create a high generalization ability and the model should be used in combination with entropy to be effective. The training set's entropy value (H) is computed in bits, as shown in Eq. (3), and may be written as:

$$H = -\sum_{i=1}^{n} P(C_i) log_2 P(C_i) \quad (3)$$

Here, H is the entropy and P is the probability for different samples.

Decision trees with excellent generalization and processing capabilities are produced by learning from unknown outcome samples. Divide and conquer is a fundamental strategy that is easy to use. The information would then be broken up into smaller groups and evaluated again until each portion has an entropy of zero. A little change in the data provided to the classifier might cause the DT structure to become exceedingly unstable. They have a lesser level of accuracy when dealing with similar types of data. If the data is connected or ambiguous, the determination of the nodes is extremely difficult. It is not appropriate for issues in which there is very little information about the data.

3.4.2. Random forest (RF) classifier

RF is a classification approach based on ensemble learning. The method employs the self-sampling strategy to build a fresh training sample set for DT by randomly extracting N samples from the old training set with a retrieval repeat. Then repeat this stage to build M more DTs at random; this is how the RF method got its name. The number of votes in the classification tree is used to determine the results of multiple decision trees. RF classification has several advantages, including improved classification accuracy, a novel technique for discovering significant variables, and the capacity to function in a complicated interaction model, including statistical data analysis and a missing value algorithm. Even huge datasets with many characteristics may be exploited using RF. It determines the value of each characteristic. It can handle imbalanced datasets and does not overfit the data. It is difficult to comprehend. Real-time categorization prediction becomes tougher as the number of trees grows. When a huge number of trees are created, the model becomes slower.

3.4.3. Naïve bays (NB) classifier

The Bayes theorem states that the probability of occurrence of one event is distinct from the other. The NB classifier is a probabilistic model based on the Bayes theorem. The approach is based on the assumption that the existence of any feature in the class is independent of the presence of any other characteristic. The possibilities of all class characteristics lead to categorizing unknown data into specified classes on their own. The Bayes theorem is used to calculate the prediction error of the class based on the probability of the predictor variable P(c/x). Class C is unaffected by the probability of any of the other variables in the class. The NB method is a scalable and fast classification technique. It may be used to solve problems with binary and multi-class categorization. It's simple to set up since it relies on basic counts of class conditional independence. With a tiny dataset, it works great. The NB technique cannot draw any link between the properties of the class since it is based on the premise that they are independent of one another. The algorithm implementation becomes more difficult when dealing with massive datasets.

3.4.4. Extra tree (ET) classifier

The ET approach uses the traditional top-down construction method to create an ensemble of unpruned decision or regression trees. It separates nodes by selecting cut points completely at random, which sets it apart from previous tree-based ensemble approaches. It also grows the trees using the entire learning sample rather than a bootstrap replica. With the initial learn-

ing sample, an ensemble model is created by repeatedly applying the ETs splitting process for numerical attributes. The final prediction is produced by combining the tree forecasts, either by arithmetic average for regression problems or by majority votes for classification issues. But because the node splitting process is so straightforward, they anticipate the constant factor to be significantly lower than in other ensemble-based approaches that prefer to utilize default cut-point settings while simultaneously locally optimizing cut-points to maximize the method's computing benefits. When a numerical value is used for the feature property, an arbitrary number between the extreme and least value is chosen at random. As a result, the sample is divided into two branches based on the characteristic property. The divergence values of all the feature traits in the node were acquired using the aforesaid approach after negotiating all of the feature attributes in the node. To achieve the node's bifurcation, choose the form with the highest bifurcation value. This procedure is more arbitrary, implying that it has a greater ability to explore.

3.4.5. XG-Boost classifier

Extreme Gradient Boosting (XG-Boost) is a popular approach for classifying huge datasets in a short period. There are several compensations to this algorithm, which contribute to its current admiration. It is a carefully scrutinized parallelized implementation of the gradient boosting technique, which allows customers to acquire their findings faster. The training time is significantly increased by parallelizing the entire boosting procedure. There are hundreds of models on diverse subsets of the training dataset instead of developing the best model feasible on the data (as in traditional approaches) and then selecting the model that performs the best. In many situations, XG-Boost outperforms conventional gradient boosting methods. A large number of core parameters are accessible in the implementation, allowing for fine-tuning for increased precision and accuracy. When compared to other classifiers, the XG-Boost classifier excelled and produced the most notable results. The classifier's parameters can also be tweaked to improve the results. XG-Boost's target function is specified in Eq. (4) given below:

$$F(\phi) = L(\phi) + \Omega(\phi) \tag{4}$$

$$\text{where } L(\phi) = l(\widehat{y}_i\, y_i) \tag{5}$$

$$\Omega(\phi) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{6}$$

Here, ϕ is a term that refers to the formula's different parameters, $L(\phi)$ is a differentiable convex loss function that trials the variance between the prediction $\widehat{y}_i$ and the target y_i as shown in Eq. (5). $\Omega(\phi)$ is a standardized term that fines complex models as shown in Eq. (6) in which T is the number of leaves in the tree, and γ is the learning rate which ranges from 0 to 1, ω_j is the weight of the leaves, and λ is a regularized parameter. Increase this item's value to further inhibit fitting and improve the model's generalization capacity. The order in which all of the FS strategies are used is crucial since the accuracy of the models is affected by picking characteristics in various combinations. Furthermore, these FS strategies are organized in such a way that the model's calculation time may be reduced. The calculation time will gradually rise if the sequence of FS techniques is changed.

3.5. Prediction and evaluation

This step utilizes a test dataset to make estimates for the model that has been created, and then cross-validation is performed. To show the findings and comparison of the classifiers, performance measures such as accuracy, precision, recall, and f-score are evaluated. Each of these evaluation parameters is obtained from the four basic features of the confusion matrix, which reflect the real and predicted groups. We can easily assess the prediction performance of each model for different types of attack activity in the training set and test set after training by looking at the matrices below. The following metrics will also be used to evaluate the efficiency of our proposed solution.

- Classification Accuracy: The percentage of properly defined data from the complete set indicated by the TP and TN condition or the percentage ratio of correctly specified data to the full dataset is known as classification accuracy, as shown in Eq. (7).

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + F_p + F_n + T_n} \tag{7}$$

Here, TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

- Precision: The uniformity of the measuring findings is referred to as precision shown in Eq. (8).

$$\text{Precision} = \frac{T_p}{T_p + F_p} \tag{8}$$

- Recall: The proportion of similar occurrences recovered is known as recall, shown in Eq. (9).

$$\text{Recall} = \frac{T_p}{T_p + F_n} \tag{9}$$

- F1-score: The weighted harmonic mean of precision and recall is the F-score shown in Eq. (10).

$$\text{F1} - \text{score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

4. Experimental results

4.1. Quantitative analysis

First, using various assessment and prediction factors, the proposed experimental analysis arrived at the results where the performance of each model in the training set and test set can be predicted clearly. The inferences achieved from the proposed FS process helped in determining the ranking of the most relevant features. They were further chosen and acted as input to the ML models, which classified them into normal or anomaly classes of traffic packets. Using our proposed approach, the FS procedure is conducted to identify the ten highest-ranking features that make up one-fourth of the NSL-KDD dataset. In this study, the training and test sets, KDDTrain.csv and KDDTest.csv, respectively, had 126,620 and 22,850 instances. Every classifier included in the model is trained many times with various combinations of features. Additionally, 8 % of the training data and 20 % of the testing data are used to validate the complete assessed model. The results produced by this approach will always be less biased, making it the most trustworthy validation technique. To produce the intended outcomes, the suggested approach has been integrated with several ML training models. The fact that NSL-KDD is open source and easily accessible online is what drives its use in our work. The outcomes of using the FCBFS approach combined with various training models are shown in Table 3.

The experiment is run on a work station (GPU) with 128 GB (127 GB usable) of RAM and Intel(R) Xeon(R) Silver 4214 CPU @ 2.20 GHz 2.19 GHz (2 processors). Following a quick correlation-based FS on the dataset, the model is trained using each classifier utilizing the top ten automatically determined features. For our experimental study, we employ Python software [40], and an ML tool composed of several ML algorithms. Python's arguments are set to their default settings during categorization. Utilizing an open-source NSL-KDD dataset, we evaluate the effectiveness of our suggested technique. The classification accuracy, precision, recall, and F-score of our technique with ten features have improved. Fig. 3 displays the performance metrics for all five classifiers using the FS approach.

To produce a significant accurate and minimal spurious item through the regular repetition of the model, the XG-Boost classifier combines a large number of tree versions with lower distinguishing reliability. Although XG-Boost uses fewer resources than current approaches, it can scale up to billions of successful samples. Furthermore, it can be approximated on the out-of-core, conserving memory on the controller. Results from five different training methods were compared, wherein it was found that FCBFS with the XG-Boost classifier was producing higher positive outcomes across the board. The classification accuracy for our proposed method with different classification techniques is also shown in Fig. 3. The qualitative analysis in the next section includes the accuracy comparison of different existing classifiers with our proposed method's results.

4.2. Qualitative analysis

The model's accuracy, precision, recall, and F1-score values are computed for all five classifiers using the FCBFS approach. This suggests that the redundant feature, which contributes next to nothing to class identification, is present to some extent in the original dataset. Comparing our suggested strategy to the entire feature set and individual filter selection methods, it was found that it takes the least amount of time to create the model. Table 4 provides a comparison of the acquired accuracy with other methods currently in use. As a result, our FS approach is effective and simple.

As presented in Table 4, the proposed technique with the best ten features presents an improvement in the classification accuracy as compared with the work in [24,36,38], and [41–45] using the NLS-KDD dataset. Following the selection of the best features using FCBFS, accuracy, precision, recall, and F1-score are obtained with ML approaches, including DT, RF, NB, ET, and XG-Boost. The findings depict how the classification algorithms may be used in conjunction with the FS approach to increase the system's accuracy and efficiency shown in Fig. 4.

The goal of implementing feature engineering is to raise the classification accuracy of the system by choosing pertinent and significant characteristics from the feature collection. The experimental results shown in quantitative and qualitative analysis depict that the combination of the XG-Boost classifier with the proposed FCBFS process yields a comparatively better result with an accuracy of 99.84 %. The results showed that our strategy, which uses 10 features, has an improvement in accuracy by 0.17 % when compared with [38], 0.71 % with [44], 0.18 % with [24], and 0.24 % with [36].

5. Conclusion and future scope

In this research, an FCBFS method in conjunction with DT, RF, NB, ET, and XG-Boost has been proposed to preprocess datasets in a heterogeneous cluster-based WSN for IoT applications. The suggested model is trained many times

Table 3 Performance metrics with different classifiers.

Classifiers	Accuracy	Precision	Recall	F1-Score
Decision Tree	99.74	99.72	99.74	99.72
Random forest	99.75	99.72	99.75	99.73
Naïve Bayes	97.52	52.46	95.70	69.64
Extra Tree	99.67	99.64	99.67	99.66
XG-Boost	99.84	99.83	99.84	99.82

Please cite this article in press as: R. Yadav et al., Augmentation in performance and security of WSNs for IoT applications using feature selection and classification techniques, Alexandria Eng. J. (2022), https://doi.org/10.1016/j.aej.2022.10.033

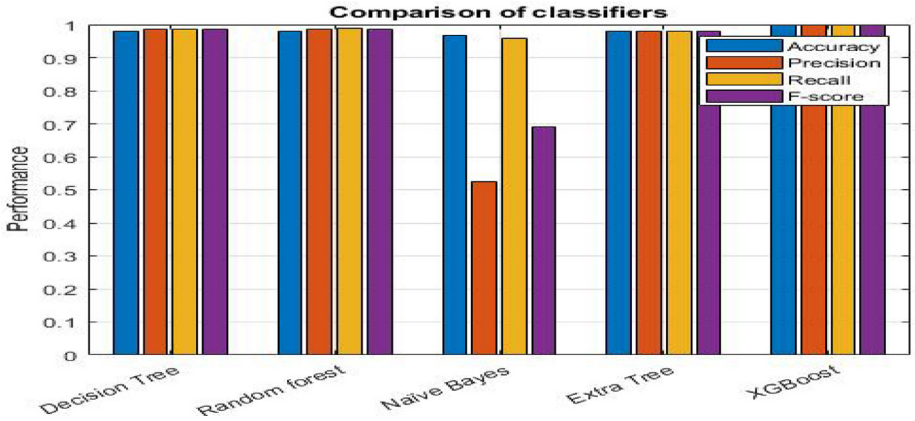


Fig. 3 Performance metrics comparison.

Table 4 Comparative analysis of proposed technique with existing methods.

Year	Author's Name	Feature Selection	Classifier	Accuracy (%)
2015	D. Deshmukh[41]	FCBFS	NB	88.20
			NB Tree	94.60
2015	B. Ingre[42]	CBFS	ANN	97.50
2016	K. Kumar[43]	CBFS	NB	97.52
		IG	NB	97.11
2016	O. Osanaiye[38]	Ensembled	DT	99.67
2017	B. Subba[44]	PCA	SVM	99.13
			DT	96.85
2018	H. Benaddi[45]	PCA	k-NN	94.00
2019	M. Amir[24]	K-mean with PCA	KNN	95.50
			RF	96.66
2021	A.Thakkar[36]	IG, Chi-square, RFE	SVM	99.60
2022	**Proposed Technique**	**Fast Correlation-based feature selection (FCBFS)**	DT	99.74
			RF	99.75
			NB	97.52
			ET	99.67
			XG-Boost	**99.84**

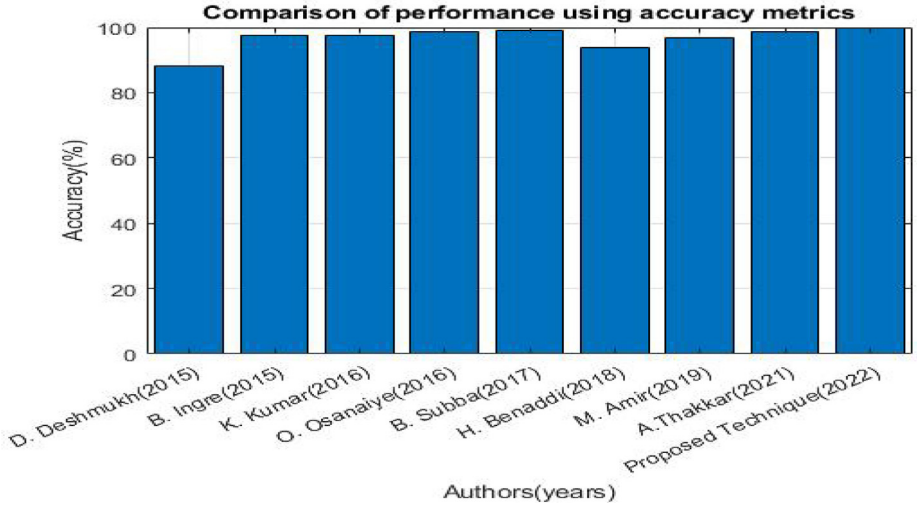


Fig. 4 Accuracy comparison with existing methods.

to achieve reproducible results using various classifiers, and the inferences are contrasted with those obtained using the current techniques. Evaluation metrics such as accuracy, precision, recall, and F1-Score have been calculated to gauge the performance and robustness of the proposed research work. A classic accuracy score of 99.84 % is achieved in the case of the XG-Boost classifier, wherein the best ten obtained features were selected after applying the proposed FCBFS. The proposed technique, which has ten features, outperforms the existing techniques in the literature for the NSL-KDD dataset. The obtained findings also show that faster correlation and other classifiers attained higher accuracy, but XG-Boost performed better overall. The proposed technique of the XG-Boost classifier with the FCBFS technique has enhanced the performance and accuracy of WSNs and their IoT applications.

In the future, first, we intend to continue our research and see how our strategy affects homogenous WSNs. One of the main issues with the NSL-KDD dataset is the computation time that can be minimized by using optimization and deep learning algorithms. The ensemble approach may be employed for training, and the nature-inspired hybrid algorithms can be utilized for optimization. Second, integration of supervised learning with unsupervised learning, in our opinion, is the best course of action to overcome the overfitting and false positives challenging issues. Third, the performance of IDS models can be enhanced by applying stack-based FS algorithms.

Funding Statement

The authors received no specific funding for this study.

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

This work has been carried out under the supervision of Prof. S. Indu, Department of ECE, Delhi Technological University, and Prof. Daya Gupta, Department of CSE, Delhi Technological University, and I (Rajiv Yadav) pay immense gratitude to my guides and UGC for enlightening me through the process.

References

[1] O.O. Ogundile, A.S. Alfa, A survey on an energy-efficient and energy-balanced routing protocol for wireless sensor networks, Sens. (Switzerland) 17 (5) (2017) 1–51, https://doi.org/10.3390/s17051084.

[2] R. Xie, X. Jia, Transmission-efficient clustering method for wireless sensor networks using compressive sensing, IEEE Trans. Parallel Distrib. Syst. 25 (3) (2014) 806–815, https://doi.org/10.1109/TPDS.2013.90.

[3] D. Fotue, F. Melakessou, H. Labiod, and T. Engel, "Mini-sink mobility with diversity-based routing in wireless sensor networks, in: PE-WASUN'11 - Proc. 8th ACM Symp. Perform. Eval. Wirel. Ad Hoc, Sensor, Ubiquitous Networks, 2011, pp. 9–16, doi: 10.1145/2069063.2069066.

[4] A. Goyal, S. Mudgal, S. Kumar, A Review on Energy-Efficient Mechanisms for Cluster-Head Selection in WSNs for IoT Application, IOP Conf. Ser. Mater. Sci. Eng. 1099 (1) (2021), https://doi.org/10.1088/1757-899x/1099/1/012010 012010.

[5] M.N.U. Islam, A. Fahmin, M.S. Hossain, M. Atiquzzaman, Denial-of-Service Attacks on Wireless Sensor Network and Defense Techniques, Wirel. Pers. Commun. 116 (3) (2021) 1993–2021, https://doi.org/10.1007/s11277-020-07776-3.

[6] G. Liu, H. Zhao, F. Fan, G. Liu, Q. Xu, S. Nazir, An Enhanced Intrusion Detection Model Based on Improved kNN in WSNs, Sensors 22 (4) (2022) 1–18, https://doi.org/10.3390/s22041407.

[7] M. Abdullah, A. Balamash, A. Al-Shannaq, S. Almabdy, Enhanced Intrusion Detection System using Feature Selection Method and Ensemble Learning Algorithms, Int. J. Comput. Sci. Inf. Secur. 16 (December) (2018) 48–55.

[8] A. Khanna, S. Kaur, Internet of Things (IoT), Applications and Challenges: A Comprehensive Review, Wireless Pers. Commun. 114 (2) (2020) 1687–1762.

[9] T.D. Diwan, S. Choubey, H.S. Hota, A Detailed Analysis on NSL-KDD Dataset using various Machine Learning Techniques for Intrusion Detection, Turkish J. Comput. Math. Educ. 12 (11) (2021) 2954–2968.

[10] K. Siddique, Z. Akhtar, F. Aslam Khan, Y. Kim, KDD Cup 99 Data Sets: A Perspective on the Role of Data Sets in Network Intrusion Detection Research, Computer (Long. Beach. Calif) 52 (2) (2019) 41–51, https://doi.org/10.1109/MC.2018.2888764.

[11] O. Osanaiye, O. Ogundile, F. Aina, A. Periola, Feature selection for intrusion detection system in a cluster-based heterogeneous wireless sensor network, Facta Univ. - Ser. Electron. Energ. 32 (2) (2019) 315–330, https://doi.org/10.2298/fuee1902315o.

[12] S. Potluri, C. Diedrich, Accelerated deep neural networks for enhanced Intrusion Detection System, IEEE Int. Conf. Emerg. Technol. Fact. Autom. ETFA, vol. 2016-Novem, pp. 1–8, 2016, doi: 10.1109/ETFA.2016.7733515.

[13] A. Zhou, Z. Li, Y. Shen, Anomaly detection of CAN bus messages using a deep neural network for autonomous vehicles, Appl. Sci. 9 (15) (2019) pp, https://doi.org/10.3390/app9153174.

[14] I. Almomani, B. Al-Kasasbeh, M. AL-Akhras, WSN-DS: A Dataset for Intrusion Detection Systems in Wireless Sensor Networks, J. Sensors 2016 (2016) 1–16.

[15] A. I. Al-Issa, M. Al-Akhras, M. S. Alsahli, M. Alawairdhi, Using machine learning to detect dos attacks in wireless sensor networks, in: 2019 IEEE Jordan Int. Jt. Conf. Electr. Eng. Inf. Technol. JEEIT 2019 - Proc., 2019, pp. 107–112, doi: 10.1109/JEEIT.2019.8717400.

[16] K. Alrawashdeh, C. Purdy, Toward an online anomaly intrusion detection system based on deep learning, in: Proc. - 2016 15th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2016, 2017, pp. 195–200, doi: 10.1109/ICMLA.2016.167.

[17] B. Ghojogh, A. Ghodsi, F. Karray, M. Crowley, Restricted Boltzmann Machine and Deep Belief Network: Tutorial and Survey, 2021, [Online]. Available: http://arxiv.org/abs/2107.12521.

[18] R. Doshi, N. Apthorpe, N. Feamster, Machine learning DDoS detection for consumer internet of things devices, in: Proc. - 2018 IEEE Symp. Secur. Priv. Work. SPW 2018, no. Ml, pp. 29–35, 2018, doi: 10.1109/SPW.2018.00013.

[19] Y. Cheng, Y. Xu, H. Zhong, Y. Liu, Leveraging Semisupervised Hierarchical Stacking Temporal Convolutional Network for Anomaly Detection in IoT Communication, IEEE Internet Things J. 8 (1) (2021) 144–155, https://doi.org/10.1109/JIOT.2020.3000771.

[20] M. Zekri, S. El Kafhali, N. Aboutabit, Y. Saadi, DDoS attack detection using machine learning techniques in cloud computing environments, in: Proc. 2017 Int. Conf. Cloud Comput. Technol. Appl. CloudTech 2017, vol. 2018-Janua, no. October, pp. 1–7, 2018, doi: 10.1109/CloudTech.2017.8284731.

[21] S. Sumathi, R. Rajesh, N. Karthikeyan, DDoS Attack Detection Using Hybrid Machine Learning Based IDS Models, J. Sci. Ind. Res. (India) 81 (3) (2022) 276–286.

[22] A. Khamparia, S. Pande, D. Gupta, A. Khanna, A.K. Sangaiah, Multi-level framework for anomaly detection in social networking, Libr. Hi Tech 38 (2) (2020) 350–366, https://doi.org/10.1108/LHT-01-2019-0023.
[23] M. Aldwairi, L. Tawalbeh, Security techniques for intelligent spam sensing and anomaly detection in online social platforms, Int. J. Electr. Comput. Eng. 10 (1) (2020) 275–287, https://doi.org/10.11591/ijece.v10i1.pp275-287.
[24] M. Aamir, S.M. Ali Zaidi, "Clustering based semi-supervised machine learning for DDoS attack classification", J. King Saud Univ. - Comput Inf. Sci. 33 (4) (2021) 436–446, https://doi.org/10.1016/j.jksuci.2019.02.003.
[25] L. Su, Y. Yao, N. Li, J. Liu, Z. Lu, B. Liu, Hierarchical Clustering Based Network Traffic Data Reduction for Improving Suspicious Flow Detection, in: Proc. - 17th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. 12th IEEE Int. Conf. Big Data Sci. Eng. Trust. 2018, pp. 744–753, 2018, doi: 10.1109/TrustCom/BigDataSE.2018.00108.
[26] S. Pande, A. Khamparia, D. Gupta, D.N.H. Thanh, DDOS Detection Using Machine Learning Technique, vol. 921. Springer Singapore, 2021.
[27] 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, ACM Int. Conf. Proceeding Ser. (2018), https://doi.org/10.1145/3167918.3167951.
[28] Y. Xue, J. Qin, Partial Connection Based on Channel Attention for Differentiable Neural Architecture Search, IEEE Trans. Ind. Informatics XX (Xx) (2022) 1–10, https://doi.org/10.1109/tii.2022.3184700.
[29] Y. Xue, Y. Tang, X. Xu, J. Liang, F. Neri, Multi-Objective Feature Selection with Missing Data in Classification, IEEE Trans. Emerg. Top. Comput. Intell. 6 (2) (2022) 355–364, https://doi.org/10.1109/TETCI.2021.3074147.
[30] Y. Xue, B. Xue, M. Zl, Self-Adaptive particle swarm optimization for large-scale feature selection in classification, ACM Trans. Knowl. Discov. Data 13 (5) (2019) pp, https://doi.org/10.1145/3340848.
[31] A. Jain, M. Khari, E. Verdú, S. Omatsu, R.G. Crespo, A route selection approach for variable data transmission in wireless sensor networks, Cluster Comput. 23 (3) (2020) 1697–1709, https://doi.org/10.1007/s10586-020-03115-0.
[32] P. Singh, M. Khari, S. Vimal, EESSMT: An Energy Efficient Hybrid Scheme for Securing Mobile Ad hoc Networks Using IoT, Wirel. Pers. Commun. 126 (3) (2022) 2149–2173.
[33] M. Rajesh, W. Wu, Recent research in network security, Int. J. Smart Sens. Intell. Syst. 2017 (Special Issue) (2017) 8–10, https://doi.org/10.21307/ijssis-2017-274.
[34] S. Vimal, M. Khari, R.G. Crespo, L. Kalaivani, N. Dey, M. Kaliappan, Energy enhancement using Multiobjective Ant colony optimization with Double Q learning algorithm for IoT based cognitive radio networks, Comput. Commun. 154 (March) (2020) 481–490, https://doi.org/10.1016/j.comcom.2020.03.004.
[35] M. Tavallaee, E. Bagheri, W. Lu, A.A. Ghorbani, A detailed analysis of the KDD CUP 99 data set, in: IEEE Symp. Comput. Intell. Secur. Def. Appl. CISDA 2009, no. Cisda, 2009, pp. 1–6, doi: 10.1109/CISDA.2009.5356528.
[36] A. Thakkar, R. Lohiya, Attack classification using feature selection techniques: a comparative study, J. Ambient Intell. Humaniz. Comput. 12 (1) (2021) 1249–1266, https://doi.org/10.1007/s12652-020-02167-9.
[37] H. Ao, Using Machine Learning Models to Detect Different Intrusion on NSL-KDD, in: 2021 IEEE Int. Conf. Comput. Sci. Artif. Intell. Electron. Eng. CSAIEE 2021, 2021, pp. 166–177, doi: 10.1109/CSAIEE54046.2021.9543241.
[38] O. Osanaiye, H. Cai, K.K.R. Choo, A. Dehghantanha, Z. Xu, M. Dlodlo, Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing, Eurasip J. Wirel. Commun. Netw. 1 (2016) 2016, https://doi.org/10.1186/s13638-016-0623-3.
[39] S. Pande, A. Khamparia, D. Gupta, Feature selection and comparison of classification algorithms for wireless sensor networks, J. Ambient Intell. Humaniz. Comput., no. 0123456789, 2021, doi: 10.1007/s12652-021-03411-6.
[40] C. Ozgur, T. Colliau, G. Rogers, Z. Hughes, MatLab vs. Python vs. R, J. Data Sci. 15 (3) (2021) 355–372, https://doi.org/10.6339/jds.201707_15(3).0001.
[41] D. H. Deshmukh, T. Ghorpade, P. Padiya, Improving classification using pre-processing and machine learning algorithms on NSL-KDD dataset, in: Proc. - 2015 Int. Conf. Commun. Inf. Comput. Technol. ICCICT 2015, pp. 13–18, 2015, doi: 10.1109/ICCICT.2015.7045674.
[42] B. Ingre, A. Yadav, Performance analysis of NSL-KDD dataset using ANN, in: Int. Conf. Signal Process. Commun. Eng. Syst. - Proc. SPACES 2015, Assoc. with IEEE, pp. 92–96, 2015, doi: 10.1109/SPACES.2015.7058223.
[43] K. Kumar, J. Singh, Network Intrusion Detection with Feature Selection Techniques using Machine-Learning Algorithms, Int. J. Comput. Appl. 150 (12) (2016) 1–13, https://doi.org/10.5120/ijca2016910764.
[44] B. Subba, S. Biswas, S. Karmakar, Enhancing performance of anomaly based intrusion detection systems through dimensionality reduction using principal component analysis, in: 2016 IEEE Int. Conf. Adv. Networks Telecommun. Syst. ANTS 2016, 2017, doi: 10.1109/ANTS.2016.7947776.
[45] H. Benaddi, K. Ibrahimi, A. Benslimane, Improving the Intrusion Detection System for NSL-KDD Dataset based on PCA-Fuzzy Clustering-KNN, in: Proc. - 2018 Int. Conf. Wirel. Networks Mob. Commun. WINCOM 2018, 2019, doi: 10.1109/WINCOM.2018.8629718.



Rajiv Yadav is currently pursuing his Ph.D. in the Department of Electronics and Communication Engineering at Delhi Technological University (DTU), Delhi, India. He did his M. Tech. (Master of Technology) in Microelectronics and VLSI Design and B.Tech. in Electronics and Communication Engineering from Kurukshetra University. He qualified UGC NET-JRF in 2015. His areas of interest are Energy Efficiency and Security of Wireless Sensor Networks.



Prof. S. Indu is a Professor in the Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, and is currently the Dean of the student welfare department there. She has done a Ph.D. in Computer Engineering from the University of Delhi and M.Tech., Post B.Tech. from the University of Kerala, Kerala. She was awarded the Best Branch Counsellor award 2013 IEEE USA. Her research interests are in the field of Computer Vision, Image Processing, and wireless sensor network. She has 3 completed sponsored research projects. She has published around 120 papers in reputed journals, national and international conferences.



Prof. Daya Gupta is a Professor in the Department of Computer Engineering, Delhi Technological University, Delhi. She has done a Ph.D. in Computer Engineering from the University of Delhi and M.Sc. (Maths), Post M.Sc. Diploma (Computers Sc.) from IIT, Delhi. She is a senior member of IEEE and a life member of CSI. Her research interests are in the field of requirement engineering, network security, ad hoc networks, and image classification in remote sensing. She has published several research papers in international journals and conferences and has chaired sessions and delivered invited talks at many national and international conferences.

Please cite this article in press as: R. Yadav et al., Augmentation in performance and security of WSNs for IoT applications using feature selection and classification techniques, Alexandria Eng. J. (2022), https://doi.org/10.1016/j.aej.2022.10.033

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/364314482

Bifurcation Analysis of Current-mode controlled CUK converter

Conference Paper · October 2022

DOI: 10.1109/ASIANCON55314.2022.9909370

CITATIONS
0

READS
5

2 authors, including:



Sudarshan K Valluru
Delhi College of Engineering (Now Delhi Technological University) Govt.of NCT Delhi
45 PUBLICATIONS **157** CITATIONS

SEE PROFILE

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

meta algorithms application to nonlinear dynamical systems View project

Power Markets View project

All content following this page was uploaded by Sudarshan K Valluru on 12 October 2022.

The user has requested enhancement of the downloaded file.

2022 2nd Asian Conference on Innovation in Technology (ASIANCON)
Pune, India. Aug 26-28, 2022

Bifurcation Analysis of Current-mode controlled CUK converter

Avinash Kumar
Control of Dynamical Systems and Computation Lab
Department of Electrical Engineering
Delhi Technological University,
Delhi, India
avinash.ips2012@gmail.com

Sudarshan K. Valluru
Control of Dynamical Systems and Computation Lab
Department of Electrical Engineering
Delhi Technological University,
Delhi, India
sudarshan_valluru@dce.ac.in

2022 2nd Asian Conference on Innovation in Technology (ASIANCON) | 978-1-6654-6851-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ASIANCON55314.2022.9909370

***Abstract*-This paper analyses the bifurcation behaviour of the CUK converter operating under continuous conduction mode. Discrete-time mapping is used to analyse the bifurcation parameters operating under a closed loop with a current-controlled scheme. to study the bifurcation diagram parameters like reference current and the Input voltage are varied by keeping others constant. It has been observed that the system loses its stability and bifurcates from 1 time-periodic to 2-time-periodic. The system then enters into chaos. These phenomena have been observed in the bifurcation diagram and hence the maximum range in which the converter can operate safely is known and is validated by time-domain waveforms and phase plots.**

Index Terms– CUK converter; chaos; bifurcation; non-linearity;

I. INTRODUCTION

In this paper, the bifurcation behaviour of the CUK converter operating in a closed loop with a current-controlled scheme has been studied. Switched-mode power converters exhibit several complex behaviours such as bifurcation, change in operating region, and chaos [1]. These behaviours can cause the converter to operate outside the boundary region, have EMI issues, cause sudden spikes, damage to the converter, sensor failure, and loads.

The wideband feedback control of the nonlinear circuit was studied by [2],[3],[4]. The dc converter is modelled in discrete form by nonlinear mapping[5],[6],[7] which was derived from intra cycle operation and the long-term cycle-to-cycle behaviour was determined by iterating the mapping.

Different numerical methods and theoretical approaches have been published like the Poincaré map, bifurcation diagram, and Lyapunov to know the boundary conditions of the converter. In this paper, the numerical method which is the discrete iterative mapping is used to determine the complex behaviour of the converter. The theoretical analysis has been validated via analytical and numerical method analysis.

This paper is divided into five different sections. The derivation of the bifurcation diagram is expressed in section II. Section III talks about the control law for feedback on the flyback converter. After that, the bifurcation behaviour and phase portraits have been plotted in section IV. Finally, section V finishes with conclusions.

II. DISCRETE ITERATIVE MAPPING

The parasitic elements have been omitted to reduce complexity. In this section, some advanced algorithms have been derived to study complex behaviour.

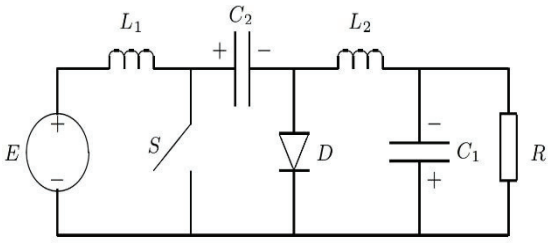


Fig.1 cuk converter schematic diagram

The state space equation for the cuk converter can be written as:

$$\dot{x} = A_1 X + B_1 E \quad \text{for } t_n \le t < t_n + dT \qquad (1)$$

$$\dot{x} = A_2 X + B_2 E \text{ for } t_n + dT \le t < t_{n+1} \qquad (2)$$

Where,

$$A_1 = \begin{bmatrix} \frac{-1}{RC_2} & 0 & \frac{1}{C_2} & 0 \\ 0 & 0 & \frac{-1}{C_1} & 0 \\ \frac{-1}{L_2} & \frac{1}{L_1} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad A_2 = \begin{bmatrix} \frac{-1}{RC_2} & 0 & \frac{1}{C_2} & 0 \\ 0 & 0 & 0 & \frac{1}{C_1} \\ \frac{-1}{L_2} & 0 & 0 & 0 \\ 0 & \frac{-1}{L_1} & 0 & 0 \end{bmatrix}$$

$$B_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{L_1} \end{bmatrix} \quad B_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{L_1} \end{bmatrix} \quad X = \begin{bmatrix} V_{c1} \\ V_{c2} \\ I_{L1} \\ I_{L2} \end{bmatrix}$$

$and\ t_{n+1} = t_n + T$

It is assumed C1 = C2 = C and L1 = L2 = L, to reduce the complexity of the differential equations. To find the general solution for the switch on interval, the RHS of (2) can be integrated and substituted with t = t_n'. After substituting, this gives the end value of switch ON interval.

$$V_{C1}(t_{n'}) = e^{\frac{-1}{RC_2}}\left(\frac{I_{L1}}{C_2}\right) + e^{\frac{-1}{RC_2}} V_{C1}(t_n) \qquad (3)$$

$$V_{C2}(t_{n'}) = \frac{-I_{L1}}{C_1}(t - t_n) + V_{C2}(t_n) \qquad (4)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 12,2022 at 08:23:59 UTC from IEEE Xplore. Restrictions apply.

$$I_{L1}(t_n') = \frac{t}{L} - \frac{t_n}{L} + I_L(t_n) \quad (5)$$

$$I_{L2}(t_n') = \frac{t-t_n}{L} + I_L(t_n) \quad (6)$$

Where, V_{C1}, V_{C2}, $I_{L1,}$ and I_{L2} are voltages across C_1, C_2, L_1 and L_2 respectively. To find the solution for the switch-off interval, apply the Laplace transformation to

$$X(s) = [SI - A_2]^{-1}\,[x(t'_n) + B_2 E(s)] \quad (7)$$

$$\begin{bmatrix} V_{C1}(s) \\ V_{C2}(s) \\ I_{L1}(S) \\ I_{L2}(S) \end{bmatrix} = \frac{1}{H(S)} \begin{bmatrix} 0 & 0 & \frac{1}{L^2} - \frac{1}{LC} + \frac{1}{L^2C} & 0 \\ 0 & 0 & 0 & \frac{1}{L^2C} \\ \frac{-1}{LC^2} & 0 & \frac{-1}{LC}(s + \frac{1}{RC}) & 0 \\ 0 & \frac{-1}{LC^2} & 0 & 0 \end{bmatrix} \begin{bmatrix} V_{C1}(t_{n'}) \\ V_{C2}(t_{n'}) \\ I_{L1}(t_n') \\ I_{L2}(t_n') \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{L} \end{bmatrix} E(s)$$

Where,

$$V_{C1}(s) = \left(\frac{\frac{1}{L^2} - \frac{1}{LC} + \frac{1}{L^2C}}{H(s)} \right) I_{L1}(t_n') \quad (8)$$

$$V_{C2}(s) = \frac{1}{L^2C}\left(\frac{I_{L2}(t_n')}{H(s)}\right) + \frac{E(s)}{L} \quad (9)$$

$$I_{L1}(S) = \frac{-V_{C1}(t_n')}{LC^2} - \frac{I_{L1}(t_n')}{LC}\left(S + \frac{1}{RC}\right) \quad (10)$$

$$I_{L2}(S) = \frac{-V_{C2}(t_n')}{LC_2} \quad (11)$$

$$H(S) = \frac{1}{LC^2} \quad (12)$$

On taking the inverse of the Laplace of the above equations 8,9,10 and 11 and equate it with 3,4,5 and 6, the resultant equations are highly complex to simulate. By taking successive approximations, putting $t_c = d_n T$ and $t_d = (1-d)T$ and ignoring the higher degree polynomial, the final matrix is as follows

$$\begin{bmatrix} v_{c1(n+1)} \\ v_{c2(n+1)} \\ v_{c3(n+1)} \\ v_{c4(n+1)} \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} & f_{13} & f_{14} \\ f_{21} & f_{22} & f_{23} & f_{24} \\ f_{31} & f_{32} & f_{33} & f_{34} \\ f_{41} & f_{42} & f_{43} & f_{44} \end{bmatrix} \begin{bmatrix} v_{c1(n)} \\ v_{c2(n)} \\ v_{c3(n)} \\ v_{c4(n)} \end{bmatrix} + \begin{bmatrix} g_{1(d_n)} \\ g_{2(d_n)} \\ g_{3(d_n)} \\ g_{4(d_n)} \end{bmatrix} E$$

$$f_{11} = \left[1 - \frac{t_d}{RC} + \frac{t^2 d}{2}\left(\frac{1}{C^2R^2} - \frac{1}{LC}\right)\right]\left[1 - \frac{t_c}{RC} + \frac{{t_c}^2}{2}\left(\frac{1}{R^2C^2} + \frac{1}{LC}\right)\right] + \left(\frac{t_d}{C} - \frac{{t_d}^2}{2C^2R}\right)\left(\frac{-t_c}{C} + \frac{{t_c}^2}{2RLC}\right) \quad (13)$$

$$f_{12} = \left[1 - \frac{t_d}{RC} + \frac{t_d^2}{2}\left(\frac{1}{C^2R^2} - \frac{1}{LC}\right)\right]\frac{t_c^2}{2LC} + \left(\frac{t_d}{C} - \frac{t_d^2}{2c^2R}\right)\frac{t_c}{L} \quad (14)$$

$$f_{13} = [1 - \frac{t_d}{RC} + t_d^2\left(\frac{1}{C^2R^2} - \frac{1}{LC}\right)\left(\frac{t_c}{C} - \frac{t_c^2}{2C^2R}\right) + \left(\frac{t_d}{C} - \frac{t_d^2}{2C^2R}\right)\left(1 - \frac{t_c^2}{LC}\right)] \quad (15)$$

$$f_{14} = 0 \quad (16)$$

$$f_{21} = \left(1 - \frac{t_d^2}{2LC}\right)\left(\frac{t_c^2}{2LC}\right) \quad (17)$$

$$f_{22} = \left(1 - \frac{t_d^2}{2LC}\right)\left(1 - \frac{t_d^2}{2LC}\right) \quad (18)$$

$$f_{23} = (1 - \frac{t_d^2}{2LC})\ (\frac{-t_c}{C}) \quad (19)$$

$$f_{24} = \frac{t_d}{C} \quad (20)$$

$$f_{31} = \left(\frac{-t_d}{L} + \frac{t_d^2}{2LCR}\right)\left[1 - \frac{t_c}{RC} + \frac{t_c^2}{2}\left(\frac{1}{C^2R^2} + \frac{d1}{LC}\right)\right] + \left(1 - \frac{t_d^2}{LC}\right)\left(\frac{-t_c}{C} + \frac{t_c^2}{2LCR}\right) \quad (21)$$

$$f_{32} = \left(\frac{-t_d}{L} + \frac{t_d^2}{2LRC}\right)\frac{t_c^2}{2LC} + \left(1 - \frac{t_d^2}{LC}\right)(1 - \frac{t_c^2}{LC}) \quad (22)$$

$$f_{33} = \left(\frac{-t_d}{L} + \frac{t_d^2}{2LRC}\right)\left(\frac{t_c}{C} - \frac{t_c^2}{2C^2R}\right) + \left(1 - \frac{t_d^2}{LC}\right)\left(1 - \frac{t_c^2}{LC}\right) \quad (23)$$

$$f_{34} = 0 \quad (24)$$

$$f_{41} = \left(\frac{-t_d}{L}\right)\left(\frac{t_c^2}{2LC}\right) \quad (25)$$

$$f_{42} = (\frac{-t_d}{L})(1 - \frac{-t_c^2}{2LC}) \quad (26)$$

$$f_{43} = \left(\frac{-t_d}{L}\right)\left(1 - \frac{-t_c}{C}\right) \quad (27)$$

$$f_{44} = 1 - \frac{t_d^2}{2LC} \quad (28)$$

$$g_1 = 0 \quad (29)$$

$$g_2 = \left(1 - \frac{t_d^2}{2LC}\right)\left(\frac{-t_d^2}{2LC}\right) + \frac{t_d^2}{LC}\left(1 - \frac{t_c^2}{LC}\right) \quad (30)$$

$$g_3 = 0 \quad (31)$$

$$g_4 = \frac{-t_d^2}{L}\left(1 - \frac{t_d^2}{2LC}\right) - \frac{t_d}{L}\left(\frac{-t_d^2}{2LC}\right) + \frac{t_c}{L}\left(1 - \frac{t_d^2}{2LC}\right) \quad (32)$$

III. FEEDBACK CONTROL LOOP

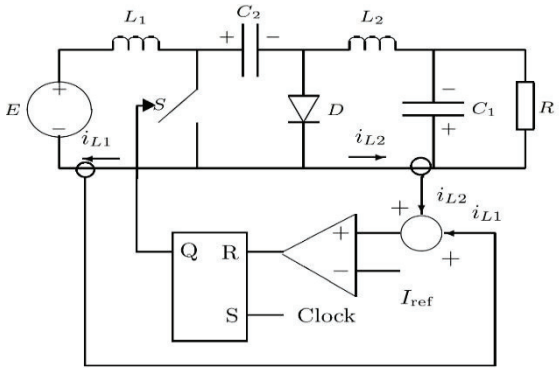


Fig.2 Current mode control of cuk converter

The cuk converter can be controlled in a number of ways. In this paper current mode control scheme was used to control output voltage. The advantage is that it has more reliability with fast, cycle current sensing for output short circuit and overload protection. So, a better study on transient can be done in current mode control. The equation for the feedback can be derived by observing the wave of open loop cuk converter.

$$I_{ref} - (i_{L1,n} - i_{L2,n}) = \left[\frac{E}{L_1} + \frac{v_{C2,n} - v_{C1,n}}{L_2}\right] d_n \quad (33)$$

$$d_n = \frac{I_{ref} - (i_{L1,n} - i_{L2,n})}{\left[\frac{E}{L_1} + \frac{v_{C2,n} - v_{C1,n}}{L_2}\right] T} \quad (34)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 12,2022 at 08:23:59 UTC from IEEE Xplore. Restrictions apply.

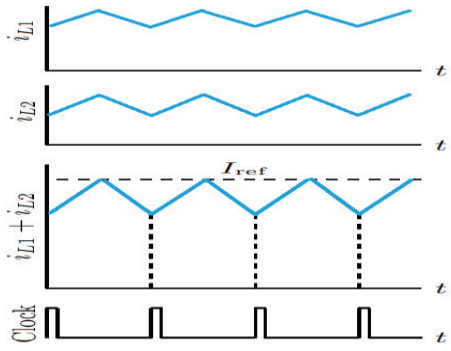


Fig.3 Current control cuk converter wave forms

IV. BIFURCATION DIAGRAMS AND PHASE POTRAITS

The bifurcation diagram is plotted to know the maximum and minimum boundary conditions in which the converter can operate safely. It is also the graphical representation from which bifurcation can be studied. For this the computer simulations are an excellent way to determine the chaotic behavior. In the computer simulation any one parameter is varied keeping others as constant.

TABLE I. EQUIPMENT VALUES

PARAMETERS	VALUES
Switching period T	200 μs
Inductance L1 and L2	16 mH
Capacitance C1 and C2	47 μF
Load resistance R	75 Ω
Input voltage E	5 V

Bifurcation diagram w.r.t Reference current

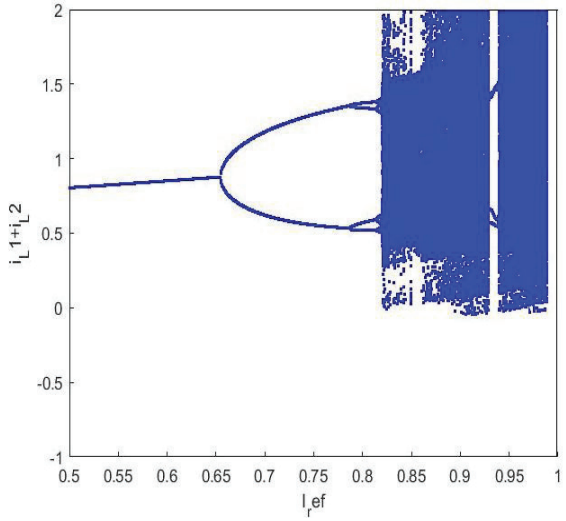


Fig.4 Bifurcation diagram w.r.t reference current Iref.

In the above Fig.4 the bifurcation diagram of sum of input current (i_{L1}) and output currents (i_{L2}) w.r.t reference current (Iref) is plotted. In this figure it is observes that the system is behaving stably up to around 0.65 A. If the Iref is increased further the system loses its stability and bifurcates and the reason is two-time attractor. If the reference current is increased further, at 0.8 A the system again bifurcates to for 4 time periodic. The system enters chaos if reference current crosses 0.83 A. the bifurcation diagram can be validated through Time domain wave forms and Phase plots.

A. Time domain wave forms

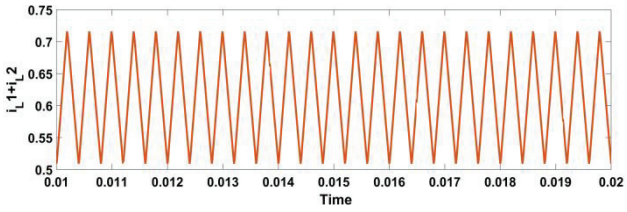


Fig.4 period-1 waveform sampled at 5 ms intervals giving one Alternating fixed points at Iref = 0.55

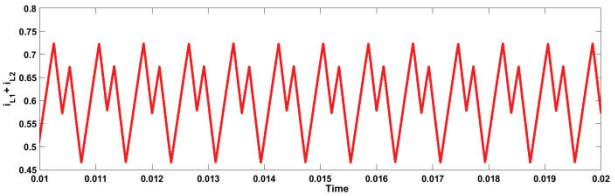


Fig.5 period-2 waveform sampled at 5 ms intervals giving one Alternating fixed points at Iref = 0.75

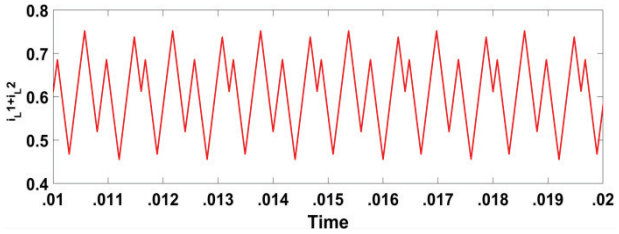


Fig.6 period-4 waveform sampled at 5 ms intervals giving one Alternating fixed points at Iref = 0.1

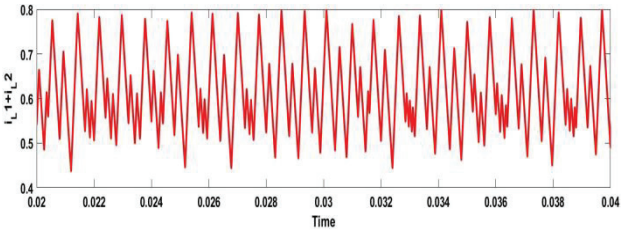


Fig.7 waveform at Chaos sampled at 5 ms intervals giving one Alternating fixed points at Iref = 0.1

The bifurcation diagram can be verified using Time domain waveforms. In this, the output current waveform has been taken. It can be observed from Fig.4 that the time domain waveform of output current repeats or mirrors itself at every 1 time period that is, the system is under stable operation. Also, Fig.5 showcases that the time domain waveform of output current repeats itself every two time periods that are the system has period-doubling bifurcations. Similarly, Fig.6 shows the time domain waveform of output current repeats itself at every 4-time period. That is, the system has 4T bifurcations. And in fig 7, the system has no period, which means the system has entered into chaos. The following section continues with the phase portraits of the bifurcation diagram.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 12,2022 at 08:23:59 UTC from IEEE Xplore. Restrictions apply.

B. Phase portraits

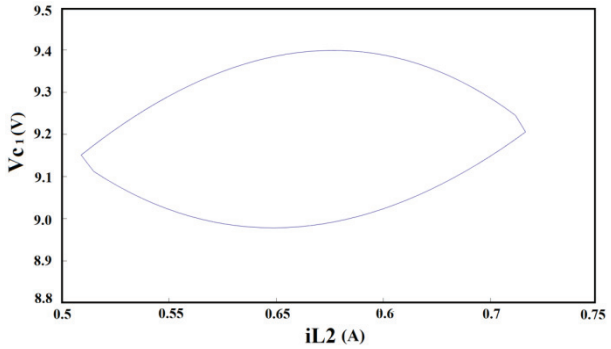


Fig .8 phase portrait of output voltage and output current at stable 1 period

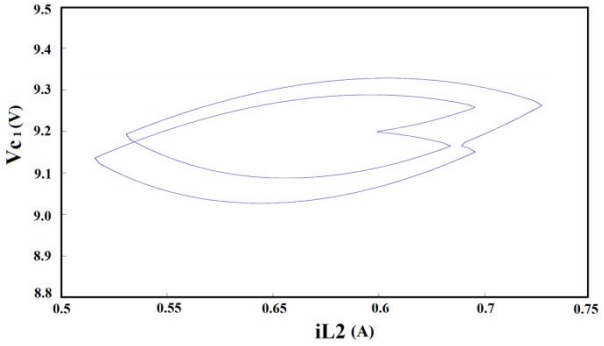


Fig .9 phase portrait of output voltage and output current at 2 time periodic.

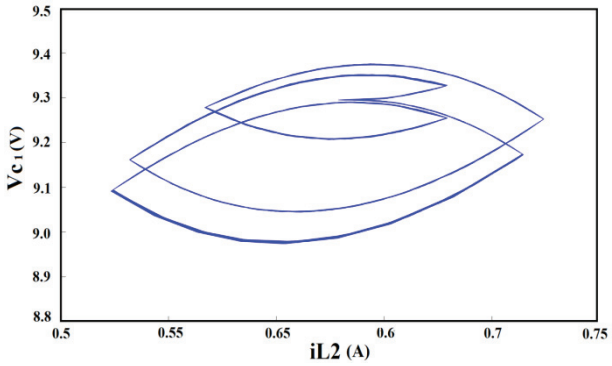


Fig .10 phase portrait of output voltage and output current at 4 time periodic

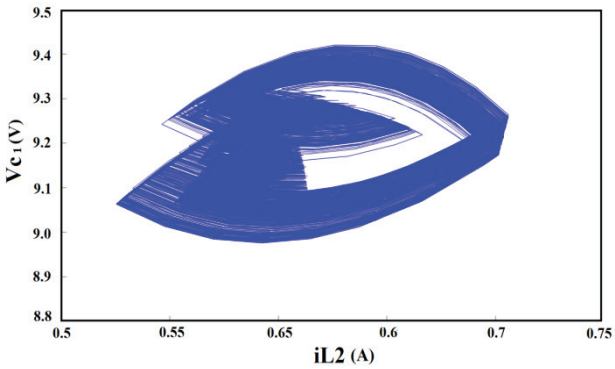


Fig .11 phase portrait of output voltage and output current at chaos

Figure 8,9,10 and 11 illustrates phase portraits of the output current vs output voltage in different stages (period 1, period 2 and chaos) of the system.

Bifurcation diagram w.r.t Load resistance

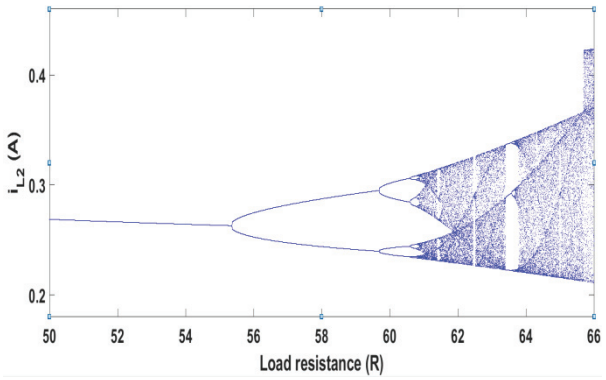


Fig.12 Bifurcation diagram w.r.t Load resistance

In figure.12 the bifurcation diagram is plotted varying load resistance. It can be observed the system remains 1T periodic stable till R= 55 ohms, there after the system loses its stability thereby causing a dynamic change and the system transforms from 1T to 2T periodic via period double bifurcations. Then the 2T periodic remains stable till R= 60 ohms. Then there after the system again loses its stability via 4T periodic. The system enters chaos after 62 ohms.

Bifurcation diagram w.r.t Input voltage

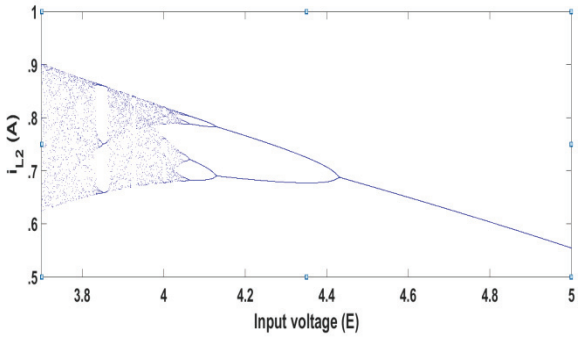


Fig.13 Bifurcation diagram w.r.t Input voltage

In the above figure.12 the bifurcation diagram is plotted varying Input voltage. It can be observed that the bifurcation happens in reverse direction. The system remains 1T periodic stable E = 4.4 V, there after the system loses its stability there by causing a dynamic change and the system transforms from 1T to 2T periodic via period double bifurcations. Then the 2T periodic remains stable till E= 4.21 V Then there after the system again loses its stability via 4T periodic. The system enters chaos after E= 4 V.

V. CONCLUSION

The research delves into the intricacies of switching power cuk converter behaviour. The associated time-domain waveforms and phase portraits have been researched to explain this complex behaviour bifurcation diagram. The next state voltage and current were calculated using discrete iterative mapping. To get the desired results, a numerical method analysis was done. Several parameters such as input voltage and feedback gain were varied and the phenomenon of bifurcation has been investigated. There have also been visual examinations of two time periods, four time periods, and pandemonium. Time-domain waveforms and phase portraits were used to verify the bifurcation diagram. The outcomes reveal the boundary conditions as well as the maximum functioning point. However, the exact bifurcation behaviour of cuk converter hasn't been studied.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 12,2022 at 08:23:59 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] C. K. Tse and M. Di Bernardo, "Complex behavior in switching power converters," *Proceedings of the IEEE*, vol. 90, no. 5, pp. 768–781, 2002.

[2] C. K. Tse, "Flip bifurcation and chaos in three-state boost switching regulators," IEEE Tran. on Circuits and Systems I, vol. 41, pp. 16-23, 1994.

[3] A. El Aroudi, G. Olivar, L. Benadero, and E. Toribio, "Hopf bifurcation and chaos from torus breakdown in PWM voltage-controlled DC-DC boost converter," IEEE Tran. on Circuits and Systems I, vol. 1 I, pp. 1374-3382,1999.

[4] D. C. Hamill, J. H. B. Deane, and D. J. Jefferies, "Modelling of the chaotic dc-dc converter by iterated mappings," IEEE Tran. on Power Electronics, vol. 7, no.9, pp. 25-33, 1992.

[5] L. G. De Vicufia, F. Guinjoan, J. Majo a L. Martinez, "Discontinuous conduction mode in the SEPlC converter," IEEE Electroteclinical Conference, MELECON'XY, pp. 38-42, 1989.

[6] M. J. Johnson, "Analysis of current-programmed SEPIC DC-DC converters in discontinuous conduction operating region," IEEE Applied Power Electronics Conference and Exposition, APEC'93, pp. 207-213, 1993.

[7] C-C. Fang and E. H. Abed, "Output regulation of dc-dc switching converters using discrete-time integral control," Proc. Of the American Control Conference, vol. 2, pp. 1052-1056, 1999, San Diego, Califomia.

[8] Sudarshan K Valluru, Aditya Verma, Anshul Gupta, "Computation of Hopf Bifurcation points in the Magnetic Levitation System," *Information and Communication Technology for Intelligent Systems. ICTIS 2020*. Smart Innovation, Systems and Technologies, vol 195. Springer, Singapore.

[9] Valluru, S.K., Singh, M., Verma, A., Gupta, "A Bifurcation curves for electrical DC motors" in 2018 *2nd IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems* (ICPEICES), Delhi, India, pp. 953–957 (2018).

[10] W. Ma, M. Wang, and C. Li, "Control of bifurcation in the one-cycle controlled Cuk converter," *Nonlinear Dynamics*, vol. 67, no. 4, pp. 2573–2583, 2012.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 12,2022 at 08:23:59 UTC from IEEE Xplore. Restrictions apply.

View publication stats

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/364314280

Bifurcation Analysis of Novel Self Operating Flyback Converter

Conference Paper · October 2022

DOI: 10.1109/ASIANCON55314.2022.9909176

CITATIONS
0

READS
19

2 authors, including:



Sudarshan K Valluru
Delhi College of Engineering (Now Delhi Technological University) Govt.of NCT Delhi
45 PUBLICATIONS **157** CITATIONS

SEE PROFILE

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

Power Markets View project

Fuzzy controller View project

All content following this page was uploaded by Sudarshan K Valluru on 12 October 2022.

The user has requested enhancement of the downloaded file.

2022 2nd Asian Conference on Innovation in Technology (ASIANCON)
Pune, India. Aug 26-28, 2022

Bifurcation Analysis of Novel Self Operating Flyback Converter

Avinash Kumar
Controls of Dynamical Systems and Computation Lab
Department of Electrical Engineering
Delhi Technological University
Delhi-110042, India
avinash.ips2012@gmail.com

Sudarshan K. Valluru
Controls of Dynamical Systems and Computation Lab
Department of Electrical Engineering
Delhi Technological University
Delhi-110042, India
-sudarshan_valluru@dce.ac.in

2022 2nd Asian Conference on Innovation in Technology (ASIANCON) | 978-1-6654-6851-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ASIANCON55314.2022.9909176

***Abstract*–This paper presents the complex behaviour of a novel self-operating flyback converter, which boosts ultra-low voltage. The flyback topology has been selected to study the complex behaviour, which is analyzed from discrete iterative mapping. The input voltage, reference current, and feedback gain parameters are validated to obtain the bifurcations. The transformation of chaotic behaviour from onetime-periodic to two time-periodic of the self-operating flyback converters is observed by varying one parameter while the other parameters are kept constant. The phase plots and respective time-domain currents were also studied. It has been observed from the bifurcation diagram at a particular point, the state of the system proceeds to the nonlinear regime.**

Index Terms– flyback, chaos, bifurcation, non-linearity, internet of things.

I. INTRODUCTION

Power electronic converter circuits exist with nonlinear switching elements, inductors, and loads, revealing a diversified and catastrophic complex behaviour during the operating regime[1]. They exhibit sudden operating regime changes, chaotic operation, infrequent instability, and sub-harmonic oscillations, which cause EMI problems and controller malfunctioning[2]. Although many studies have been conducted on bifurcation and chaotic behaviour on the buck[3], [4], boost[5], and Cuk[6] converters through iterative mapping. But discrete-time mapping is rarely concluded. So, this paper focuses on the discrete-time mapping for bifurcation and chaotic study, which is more accurate than iterative mapping.

The conventional treatment is to find ways to eliminate any complexity, often in some expeditious and heuristic manner, such as adjusting circuit components and parameters through a tribulation-and-error procedure.

The Flyback converter has the disadvantage that it cannot boost very low voltage, which is overcome by using a boost-flyback topology[7]. In this paper, the flyback topology was selected from the boost-flyback topology[8], and the feedback of the flyback converter is controlled under the peak current mode control scheme[9]. The converter is running under continuous mode

In the bifurcation study, the first step is to derive the discrete-time mapping as expressed in section II. The feedback control law is derived in section III. The bifurcation behaviour and phase portraits have been plotted in section IV. Finally, section V finishes with conclusions.

II. DISCRETE ITERATIVE MAP DERIVATION

The discrete iterative map technique is used for analysing bifurcation behaviour. All the parasitic elements in Flyback Converter shown in Fig.1 are assumed to be zero to simplify the calculations.

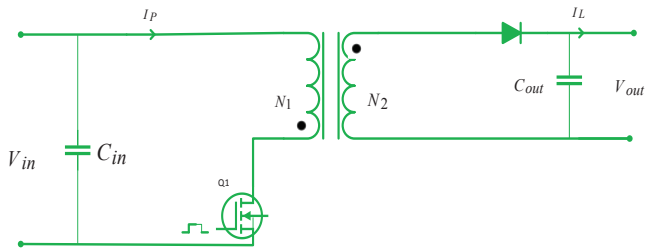


Fig.1 Flyback converter schematic diagram

The discrete iterative map for any Power electronic converter should be mapped as equation (1).

$$x_{n+1} = f(x_n) \tag{1}$$

The switching converter operates in continuous conduction mode; its topological sequence consists of two linear circuits described by the state equations (3) and (4).$\dot{x} = A_1 X + B_1 E \quad \text{for } t_n \le t < t_n + dT$ (2)

$$\dot{x} = A_2 X + B_2 E \quad \text{for } t_n + dT \le t < t_{n+1} \tag{3}$$

Where,

$$X = \begin{bmatrix} V_c \\ I_L \end{bmatrix} \quad \text{and} \quad t_{n+1} = t_n + T \tag{4}$$

$$A_1 = \begin{bmatrix} \frac{-1}{RC} & 0 \\ 0 & 0 \end{bmatrix} \qquad B_1 = \begin{bmatrix} 0 \\ \frac{1}{L} \end{bmatrix} \tag{5}$$

$$A_2 = \begin{bmatrix} \frac{-1}{RC} & (n)\left(\frac{1}{C}\right) \\ (-n)\left(\frac{1}{L}\right) & 0 \end{bmatrix} \quad B_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{6}$$

Where $n = \left(\frac{N_1}{N_2}\right)$

From Fig.2 Approximated switch ON time can be calculated by

$$t_{on} = \frac{L}{n\,V_{in}}\left(I_{ref} - I_L(t_n)\right) \tag{7}$$

The general solution for the switch-on interval to be found easily by directly integrating the RHS of (2) and putting $t = t_n'$ gives the value of $\boldsymbol{x}$ at the end of the switch-ON interval.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 12,2022 at 08:26:50 UTC from IEEE Xplore. Restrictions apply.

$$\mathrm{Vc}(\mathrm{t_n}') = \mathrm{V_c}(\mathrm{t_n})\mathrm{e}^{-(\mathrm{d_nT})/\mathrm{CR}} \quad (8)$$

$$\mathrm{I_l}(\mathrm{t_n}') = \mathrm{I_L}(\mathrm{t_n}) + \frac{\mathrm{E(dT)}}{\mathrm{L}} \quad (9)$$

Where, d_n is duty cycle at n_{th} cycle.

To find the solution for the switch-off interval, we apply the Laplace transformation to

$$\mathrm{X(s)} = [\mathrm{SI} - A_2]^{-1}[\mathrm{x}(t'_n) + B_2\mathrm{E(s)}] \quad (10)$$

$$\mathrm{X(s)} = \begin{bmatrix} s & n/c \\ -n/L & s+1/RC \end{bmatrix} \begin{bmatrix} V_c(t'_n) \\ i_L(t'_n) \end{bmatrix} \Big/ H(s) \quad (11)$$

$$\text{Where } \mathrm{H(s)} = s^2 + \frac{s}{RC} + \frac{n^2}{LC} \quad (12)$$

The s-domain expressions for the capacitor voltage and the inductor current can be written as

$$V_c\ (\mathrm{s}) = \frac{s\,V_c(t'_n) + \frac{n}{c} i_L(t'_n))}{\mathrm{H(s)}} \quad (13)$$

$$\mathrm{I_L}\ (\mathrm{s}) = \frac{\frac{-n}{L}(t'_n) + (s + \frac{1}{RC}) i_L(t'_n))}{\mathrm{H(s)}} \quad (14)$$

On solving quadratic equation H(s), the roots are,

$$\lambda_1, \lambda_2 = \frac{-1}{2RC} \pm \sqrt{\left(\frac{1}{2RC}\right)^2 - \frac{n^2}{LC}} \quad (15)$$

CASE 1 : consider system to be underdamped

$$\frac{n^2}{LC} > \left(\frac{1}{2RC}\right)^2$$

$$\lambda_1, \lambda_2 = -\sigma \pm j\omega \quad (16)$$

$$\text{Where } \omega = \frac{n^2}{LC} - \left(\frac{1}{2RC}\right)^2 \text{ and } \sigma = \frac{1}{2RC} \quad (17)$$

On applying Inverse Laplace, solving eq. 12 and 13, and putting and $t_{n+1} - t'_n = (1 - \mathrm{d}_n)\mathrm{T}$, we got the solution that was far too complex to get any result. Applying a second-degree series approximation as T/RC and σT are usually considered small.

$$\begin{bmatrix} v_c(t_{n+1}) \\ i_L(t_{n+1}) \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} \begin{bmatrix} v_c(t_n) \\ i_L(t_n) \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} E \quad (18)$$

$$f_{11} = 1 - \frac{(2-\mathrm{d_n})\mathrm{T}}{\mathrm{RC}} + \frac{1}{2}\left|\frac{\mathrm{T}}{\mathrm{RC}}\right|^2 \quad (19)$$

$$\mathrm{f_{12}} = \frac{\mathrm{n}(1-\mathrm{d_n})\mathrm{T}}{\mathrm{C}} - \frac{\mathrm{nR}}{2}\left[\frac{(1-\mathrm{d_n})\mathrm{T}}{\mathrm{RC}}\right]^2 \quad (20)$$

$$f_{21} = \frac{n(1-d_n)t}{C} + \frac{n(1-d_n^2)T^2}{2LCR} \quad (21)$$

$$f_{22} = 1 - \frac{(1-\mathrm{d_n})\mathrm{T}}{\mathrm{RC}} \quad (22)$$

$$\mathrm{g_1} = \frac{\mathrm{n\ d_nT(1-d_n)T}}{\mathrm{LC}} \quad (23)$$

$$\mathrm{g_2} = \frac{\mathrm{d_nT}}{\mathrm{L}} - \frac{\mathrm{d_nT(1-d_n)T}}{\mathrm{RLC}} \quad (24)$$

CASE 2: consider system to be undamped

$$\frac{n^2}{LC} = \left(\frac{1}{2RC}\right)^2$$

The approximate solution is

$$\begin{bmatrix} v_c(t_{n+1}) \\ i_L(t_{n+1}) \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} \begin{bmatrix} v_c(t_n) \\ i_L(t_n) \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} E \quad (25)$$

$$\mathrm{f_{11}} = \left(1 - \frac{\mathrm{T}}{\mathrm{RC}}\right)\left(1 - \frac{(1-\mathrm{d_n})^2\,\mathrm{T}^2}{2\mathrm{RC}}\right) \quad (26)$$

$$f_{12} = \frac{n}{C}\left(1 - \frac{(1-d_n)T}{2RC}\right) \quad (27)$$

$$\mathrm{f_{21}} = \frac{-\mathrm{n}}{\mathrm{L}}\left(1 - \frac{\mathrm{T}}{\mathrm{RC}} + \frac{1}{2}\left(\frac{\mathrm{T}}{\mathrm{RC}}\right)^2 - \frac{(1-\mathrm{d_n}^2)\mathrm{T}^2}{2\mathrm{LC}}\right) \quad (28)$$

$$f_{22}=1 \quad (29)$$

$$g_1 = \frac{n}{C}\frac{d_nT}{L}\left(1 - \frac{(1-d_n)T}{2RC}\right) \quad (30)$$

$$g_2 = \frac{d_nT}{L} \quad (31)$$

CASE 3: consider system to be overdamped

$$\frac{n^2}{LC} < \left(\frac{1}{2RC}\right)^2$$

The approximate solution is

$$\begin{bmatrix} v_c(t_{n+1}) \\ i_L(t_{n+1}) \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} \begin{bmatrix} v_c(t_n) \\ i_L(t_n) \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} E \quad (32)$$

$$f_{11} = \left\{\left(1 - \frac{T}{RC}\right)\cosh\ \omega(1-d)\,T - \frac{\sigma}{\omega}\,\frac{(1-d)T}{RC}\sinh\ \omega(1-d)T\right\} \quad (33)$$

$$f_{12} = \frac{n}{wc} - \frac{(1-d)T}{C}\sinh\ \omega(1-d)T \quad (34)$$

$$f_{21} = \frac{-n}{wL}\left(1 - \frac{T}{RC}\right)\sinh\ \omega(1-d)T \quad (35)$$

$$f_{22} = \left(\frac{(1-d)T}{2RC}\right)\left(\cosh\ \omega(1-d)\,T - \frac{\sigma}{\omega}\sinh\omega(1-d)T\right) \quad (36)$$

$$g_1 = \frac{n}{\omega C}\frac{dT}{L}\frac{(1-d)T}{RC}\sinh\omega(1-d)T \quad (37)$$

$$g_2 = \frac{1}{\omega\mathrm{RC}}\frac{dT}{L}\left(e^{-\sigma(1-d)T}\ \sinh\omega(1-d)T\right) \quad (38)$$

However, the most crucial parameter is case (1), which is almost equivalent to experimental conditions. So, the study of bifurcation will be based on this condition. This section has successfully provided the necessary equations and the steps to be followed to obtain the bifurcation diagram. The nest section enriches the control law for the feedback.

III. CONTROL LAW FOR FEEDBACK

A feedback mechanism is required to control the output voltage. In this paper, the feedback mechanism has been varied proportionally to reduce the complexity. The control law can be written as equation (39).

$$d_n = \mathrm{H}\ (\mathrm{D} - \mathrm{k}\ (v_{ref} - V_c\ (t_n))) \quad (39)$$

Where, D = duty cycle at steady state

k = feedback gain

v_c = capacitor voltage

$$H(x) = \begin{cases} 0\ for\ x = 0 \\ 1\ for\ x > 1 \\ x\ for\ else \end{cases}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 12,2022 at 08:26:50 UTC from IEEE Xplore. Restrictions apply.

IV. BIFURCATION BEHAVIOUR AND PHASE PORTRAITS

The bifurcation diagram is a graphical representation method generally used to study nonlinear phenomena for a system with varying parameters. In this approach, a chosen parameter is varied and plotted along with one of the axes. The monitoring state variables are sampled and plotted on the other axis as discrete points.

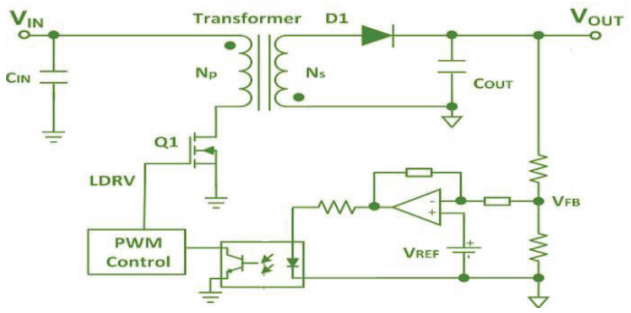


Fig2 closed-loop voltage control of flyback converter

TABLE I. CIRCUIT DETAILS

Parameters	Values
Vin	1 V
Inductor L	14.49 mH
Capacitor C	2 μF
Resistor R	2 Ohm
Switching Time T	330 μ sec

The following bifurcation diagram below has been obtained by varying feedback gain and input voltage. The output accepted can be verified using phase plots and time-domain waveforms.

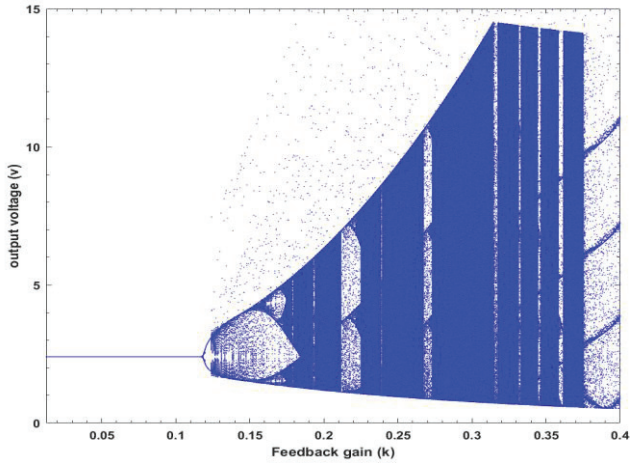


Fig.3 bifurcation diagram with varying feedback gain

In Fig.3, the bifurcation diagram regarding feedback gain can be seen, and a stable period up to almost 0.12 can be observed. After k= 0.12, there occurs a period of double bifurcation or 2T. This occurs up to k= 0.16. And after k = 0.16, the output voltage bifurcates to period 4 or 4T, which continues to almost k = 0.175. This then finally enters into a chaotic region after around k = 0.19.

A. Time-domain waveforms

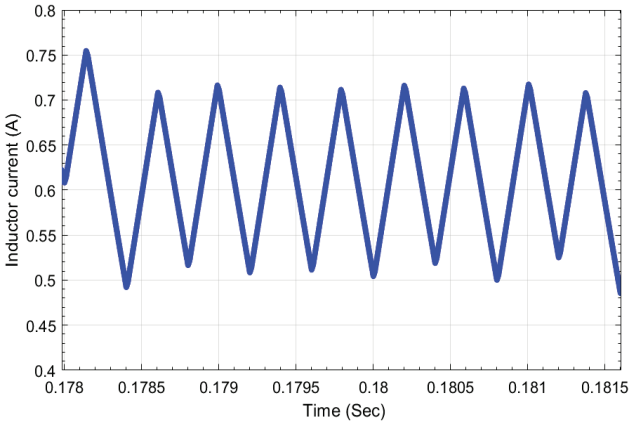


Fig.4 period-1 waveform sampled at 5ms intervals giving one Alternating fixed points at k = 0.1

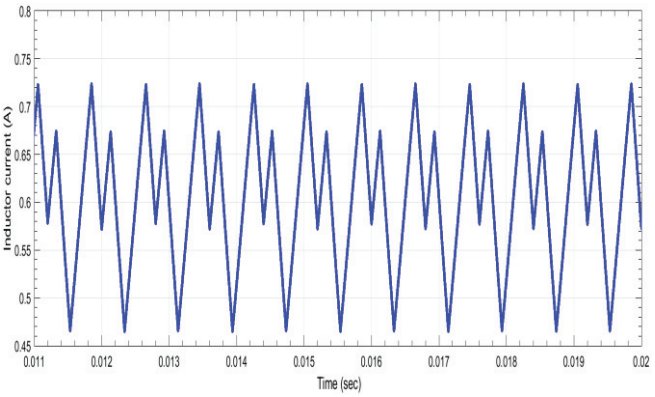


Fig.5 period-2 waveform sampled at 5 ms intervals giving two Alternating fixed points k = 0.14

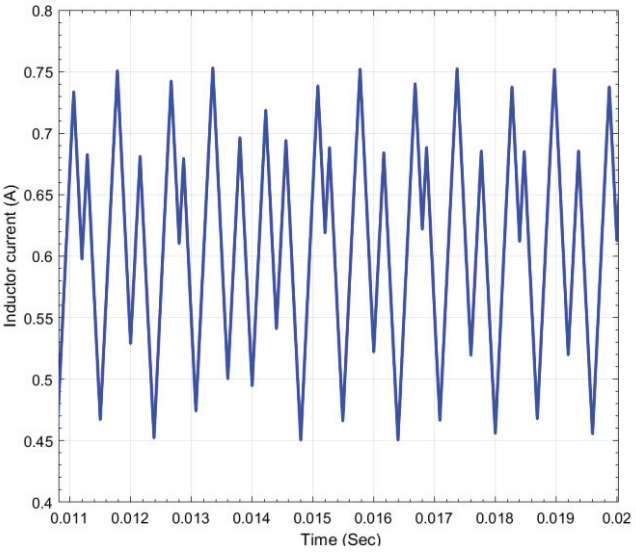


Fig.6 period-4 waveform sampled at 5 ms intervals giving four Alternating fixed points k = 0.17

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 12,2022 at 08:26:50 UTC from IEEE Xplore. Restrictions apply.

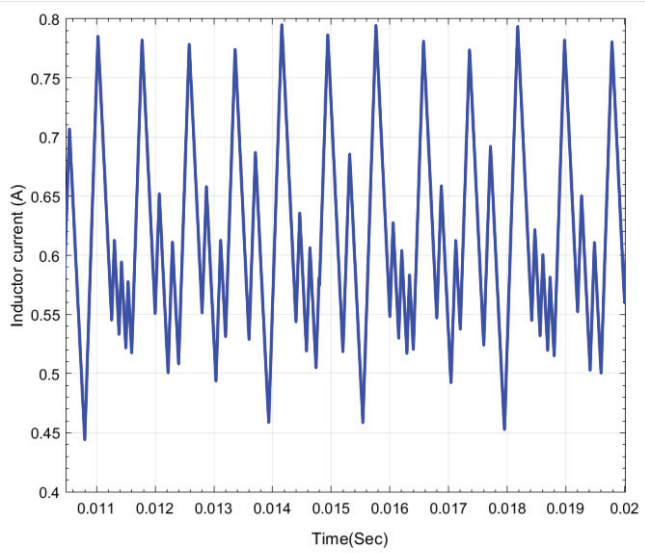


Fig.7 waveform at chaos sampled at 5 ms intervals giving no Alternating fixed points k = 0.2

The bifurcation diagram can be verified using Time domain waveforms. In this, the output current waveform has been taken. It can be observed from Fig.4 that the time domain waveform of output current repeats or mirrors itself at every 1-time period; that is, the system is under stable operation. Also, Fig.5 showcases that the time domain waveform of output current repeats itself every two time periods that the system has period-doubling bifurcations. Similarly, Fig.6 shows the time domain waveform of output current repeats itself at every 4-time period. That is, the system has 4T bifurcations. And in Fig.7, the system has no period, which means the system has entered into chaos. The following section continues with the phase portraits of the bifurcation diagram

B. Phase plots

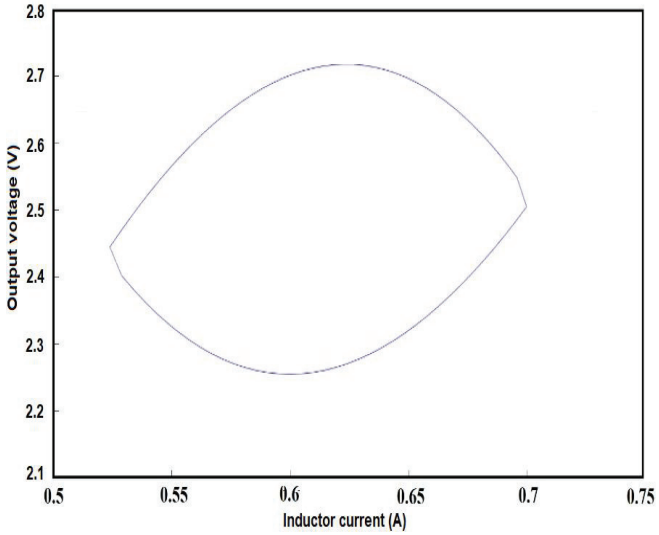


Fig .8 phase portrait of output voltage and output current at stable 1 Period

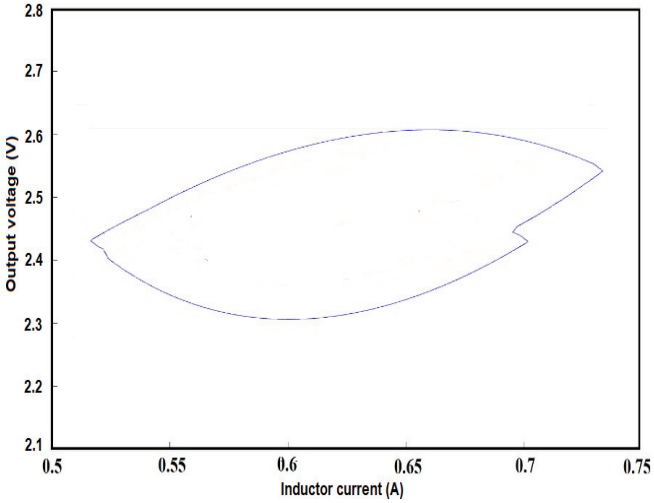


Fig .9 phase portrait of output voltage and output current at period 2.

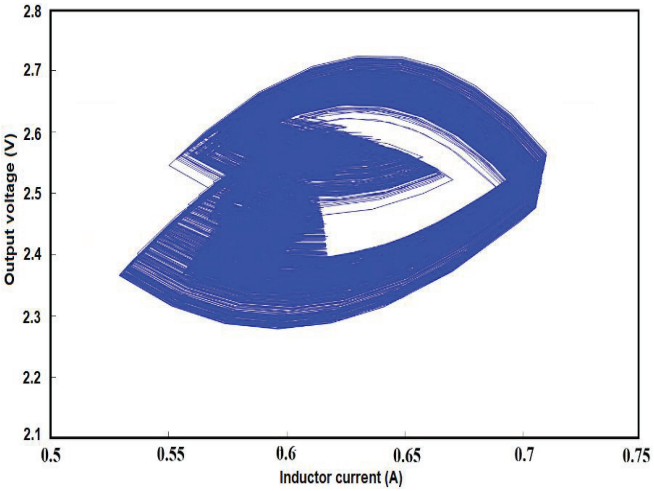


Fig .10 phase portrait of output voltage and output current at the chaos

Fig.8, Fig.9 and Fig.10 illustrate phase portraits of the output voltage in different stages (period 1, period 2 and chaos) of the system.

1) Bifurcation concerning the input voltage

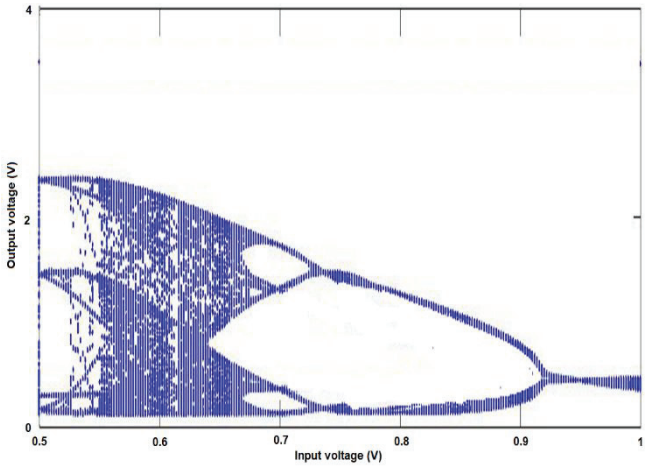


Fig .11 bifurcation diagram with varying input voltage

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 12,2022 at 08:26:50 UTC from IEEE Xplore. Restrictions apply.

In Fig.11, the bifurcation diagram about input voltage can be seen. It has been observed that the bifurcation occurs in a reverse manner. A stable one period is observed up to almost 0.92 volts. Before 0.92 volts, there arises a period of double bifurcation or 2T. This continues up to about 0.72 Volts. The output voltage again bifurcates to period 4 or 4T before 0.72 volts to almost 0.68 volts. Finally, it enters into a chaotic region before about V= 0.68. The time-domain waveforms and phase plots can also be studied in similar manners.

2) Bifurcation w.r.t reference current

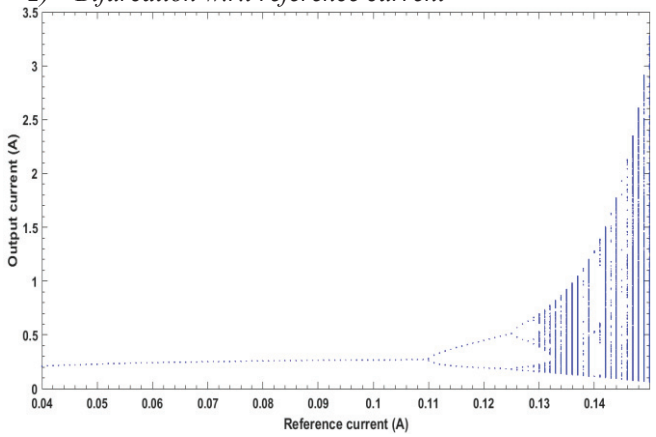


Fig.12 bifurcation diagram with varying input voltage

In Fig.12, the bifurcation diagram of the reference current is observed. A stable one period up to almost 0.11 amp is marked. After A= 0.11, there occurs a period of double bifurcation or 2T. After 0.13A, the output current bifurcates to 4T. Similarly, after 0.135A, the output current shows chaotic behaviour.

V. CONCLUSION

The paper gives an insight into the complex behaviour of switching power flyback converters. To analyze this complex behaviour bifurcation diagram, corresponding time-domain waveforms and phase portraits have been studied. Bifurcation is the most robust method of learning chaos. Discrete iterative mapping was derived to get the next state voltage and current. A numerical method analysis has been carried out to obtain the desired results. Various methods like input voltage and feedback gain were varied in a closed-loop converter with a voltage control loop. Bifurcation has been studied. Also, two time periods, four-time periods, and chaos have been visually examined. The bifurcation diagram has been verified with Time-domain waveforms and phase portraits. The results provide information about the boundary conditions and maximum operating point.

REFERENCES

[1] C. K. Tse and M. Di Bernardo, "Complex behavior in switching power converters," *Proceedings of the IEEE*, vol. 90, no. 5, pp. 768–781, 2002.

[2] David C.Hamill and David J.jeffries, "Subharmonics and Chaos in a Controlled Switched-Mode Power Converter," *IEEE Tranactions on Circuits and Systems*, vol. 35, no. 8, pp. 1059–1061, 1984.

[3] E. and G. O. Fossas, "Study of chaos in the buck converter," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 43, no. 1, pp. 13–25, 1996.

[4] J. D. Morcillo Bastidas, J. G. Munoz, and G. Olivar Tost, "Simulation and Analysis of Complex Behaviors in a DC-DC Buck Converter," *IEEE Latin America Transactions*, vol. 19, no. 1, pp. 68–74, 2021.

[5] S. Banerjee and K. Chakrabarty, "Nonlinear modeling and bifurcations in the boost converter," *IEEE Transactions on Power Electronics*, vol. 13, no. 2, pp. 252–260, 1998.

[6] W. Ma, M. Wang, and C. Li, "Control of bifurcation in the one-cycle controlled Cuk converter," *Nonlinear Dynamics*, vol. 67, no. 4, pp. 2573–2583, 2012.

[7] S. Patra *et al.*, "Self-operating Flyback Converter for Boosting Ultralow Voltage of Thermoelectric Power Generator for IoT Applications," *IEEE Transactions on Industrial Electronics*, 2021.

[8] L. L. Xie and R. X. Gong, "Study of The Mechanism of Period-doubling Bifurcation in DCM Flyback Converter," in *2014 International Power Electronics and Application Conference and Exposition*, 2014, pp. 1479–1485.

[9] J. G. Muñoz, G. Gallo, F. Angulo, and G. Osorio, "Coexistence of solutions in a boost-flyback converter with current mode control," in *8th IEEE Latin American Symposium on Circuits and Systems*, 2017, pp. 1–4.

[10] Sudarshan K Valluru, Aditya Verma, Anshul Gupta, "Computation of Hopf Bifurcation points in the Magnetic Levitation System," *Information and Communication Technology for Intelligent Systems. ICTIS 2020.* Smart Innovation, Systems and Technologies, vol 195. Springer, Singapore.

[11] Valluru, S.K., Singh, M., Verma, A., Gupta, "A Bifurcation curves for electrical DC motors" in 2018 *2nd IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems* (ICPEICES), Delhi, India, pp. 953–957 (2018).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 12,2022 at 08:26:50 UTC from IEEE Xplore. Restrictions apply.
View publication stats

2022 2nd Asian Conference on Innovation in Technology (ASIANCON)
Pune, India. Aug 26-28, 2022

Biomedical Implantable Antenna with Fractal Geometry

Nancy
Department of Electronics and Communication engineering
Delhi Technological University
New Delhi, India
nancy_2k20moc04@dtu.ac.in

Kaustubh Ranjan Singh
Department of Electronics and Communication engineering
Delhi Technological University
New Delhi, India
kaustubh@dtu.ac.in

2022 2nd Asian Conference on Innovation in Technology (ASIANCON) | 978-1-6654-6851-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ASIANCON55314.2022.9909028

***Abstract*— This paper aims at presenting the design and implementation of the fractal antenna with a dual-ring slot structure for biomedical applications in the ISM frequency band (2.4 GHz - 2.48 GHz) . Inscribed square circular fractal geometry is used to make the antenna compact and with wideband characteristics. Thin and biocompatible substrate-superstrate layers are used in designing this implantable antenna to attain human body insulation and flexibility. The human body proves to be a hostile environment from radio wave propagation in contrast with the conventional wireless system. This environment is extremely lossy and affects the antenna parameters and hence lowers the gain of the implanted antenna. To minimize this effect, techniques like encapsulation are used. Moreover, a ring and a hemispherical lens is used externally in order to enhance the gain and reduce SAR. A comparative analysis has been done for the various stages of the designed antenna to get the best one suitable for the application.**

Keywords: Fractals, Fractal geometries, biomedical applications, Gain enhancement, wideband.

I. INTRODUCTION

Wireless medical applications have made a tremendous growth in the healthcare sector over the past decade. Implantable antennas are used in heart pacemakers, retinal prosthesis and even in endoscopy in case of ulcers, cancerous tumours, inflammatory bowels etc. This technology is of great use in continuously monitoring the patient that will reduce the risk of complications. This can also be cost effective as the patient need not to be admitted in hospital in some cases. With antenna implanted in the body one can continuously monitors patients' glucose level, blood pressure, pulse rate etc and this vital information is transmitted to external devices via a wireless communication link [8].

Biomedical implantable antenna requires some basic properties like compact size, low specific absorption rate, flexible and with wideband characteristics. For proper positioning of the antenna inside the human body, the size of the antenna must be small. Various techniques are used for miniaturization of antenna including concentric ring slots in the shape of rectangle making the antenna compact and simple. Two resonant frequencies that are nearly spaced are excited by this dual ring topology that results in larger bandwidth [2]. Along with this, fractals is also used that are the self-similar patterns to maximize the effective length of the antenna which makes the antenna multiband, wideband, and also miniaturized without any use of inductors or capacitors. They are also known as space-filling curves as we can increase the perimeter while keeping its area constant. Many different fractal geometries like Sierpinski gasket, Sierpinski carpet, Koch curve, Minkowski island, etc. are used which first involves a basic shape that is repeated over and over with every iteration making it a multiband antenna [1].

Flexibility is another requirement for the implantable antenna. This prevents the breakage of the antenna with any movement of the body part or the muscle. Thin and flexible layers of substrate and superstrate are used to provide flexibility. Also, these layers should be biocompatible [4] [5]. The Antenna is covered with a thin layer of biocompatible material to enhance its performance by reducing the absorption in the lossy environment. A reduced antenna gain has been observed due to the absorbing nature of the human body. So, to improve the antenna gain, different techniques are used namely external parasitic ring and hemispherical lens along with the encapsulation layer around the antenna [2].

II. FRACTAL GEOMETRY

The main concept behind the fractal antenna was to increases the perimeter or the length of the conductor within a restricted given area or volume, that receives or radiates the electromagnetic radiations. To implement this, self-similar structures were used to design an antenna. They are also known as space-filling curves. The different iterations in a design can produce different size geometries that have different electrical length and thus resonates at different frequencies which also make these antennas wideband or even multiband [1]. Fractal design has commonly two dimensions and components:

- The Initiator (0th stage) which is the basic structure of the fractal geometry.
- The Generator which is the structure which is repeated in a sequential pattern on the initiator in the consequent stages with the different dimensions as shown in Figure.1Formula for self-similarity dimension is given as:

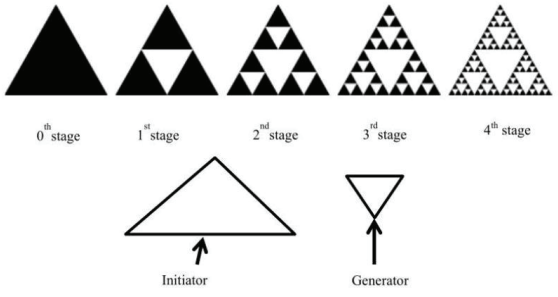


Fig.1 Sierpinski Gasket Fractal Geometry

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:08:57 UTC from IEEE Xplore. Restrictions apply.

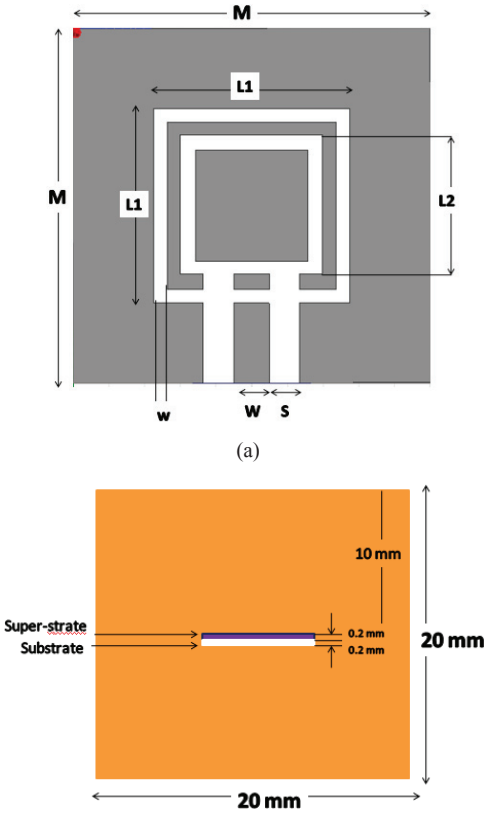
$$Ds = \log(N)/\log(1/s) \quad (1)$$

Where, `s' is the scaling factor and `N' is the number of self-similar copies.

III. ANTENNA DESIGN

A. Rectangular slots antenna

The biocompatible implantable antenna configuration comprises two rectangular slots of equal width w, shown in Fig. 2(a). These slots with the outer dimensions as L1, and inner dimensions as L2 are fabricated on a 10 mm × 10 mm polyethylene substrate with dielectric constant (εr) 2.25 and loss tangent 0.001. A superstrate layer of alumina having high a dielectric constant of 9.4 and loss tangent 0.006 is placed over the antenna. High values of the dielectric constant of the superstrate material stabilize fluctuations around the antenna due to effective permittivity [2]. It also decouples the antenna from the lossy surrounding. Substrate and superstrate materials are chosen based on impedance matching, minimum reflection, and biocompatibility. Keeping in mind the flexibility, both superstrate and substrate thickness is kept at 0.2 mm. The Total dimension of the proposed antenna is 10 mm × 10 Here, the Co-Planar Waveguide (CPW) feed is used for 50 Ω impedance matching by using appropriate trace width (W) and ground plane spacing (S). Fig. 2(b) shows the placement of the proposed antenna inside a single-layer human-muscle model. Table I shows all the design parameters of the implantable slot antenna.



(a)

(b)

Fig.2

TABLE I. PARAMETRIC VALUES OF THE PROPOSED ANTENNA.

S. No.	Parameters	Description	Values
1.	M	Width and length of substrate	10 mm
2.	L1	Outer rectangular slot	5.5 mm
3.	L2	Inner rectangular slot	4 mm
4.	W	Trace width	0.5 mm
5.	S	Ground plane spacing	1 mm
6.	w	Width of the slot	0.85 mm

B. Fractal antenna design

Considering, antenna in fig.2 (a) as 0th iteration, we will further design the fractal antenna using Inscribed square circular fractal geometry [6]. In the first iteration as shown in Figure.3 (a), ten circles of radius 1 mm are removed from three sides creating slots. One more circle of radius 1.4 mm is removed from the centre. Next step is to make squares of 1.6 mm dimension at the center of each circle and of 2.4 mm at the center circle. This completes our first iteration. Similarly, further iterations are done as shown in figure.3 (b) and (c) and the dimensions for the same are given on Table II.

IV. GAIN ENHANCEMENT OF THE IMPLANTABLE SLOT ANTENNA.

A. Encapsulation of the antenna

Casing the antenna with a dielectric material coating whose permittivity is lower than the human body tissue can significantly improve the antenna performance. This is because of the decoupling with the surrounding lossy environment which improves the gain of the antenna. Here, a thin coating of Teflon with dielectric constant 2.08 is applied around the antenna [2] [7].

B. External Parasitic Ring

By coupling the exiting electromagnetic wave to an external parasitic ring, antenna performance can be improved [2]. This ring is placed on the skin with radius 24mm and thickness 3 mm as shown in the figure.4 (a). It has improved gain by 5.2 dB.

C. External Hemispherical Lens

To minimize the spreading of incident divergent energy in undesired directions, Lens antennas are used. They collimate incident energy-producing highly directive beams by converting divergent energy waves to plane waves. The selection of material that is both mechanically and electromagnetically stable for effective lens antennas is very important [2]. The material can be selected depending on relative permittivity varying from about 1.2 to 13. To improve the implanted antenna performance an external hemispherical lens was designed and investigated to be placed on the skin over the implanted antenna as shown in the figure.4

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:08:57 UTC from IEEE Xplore. Restrictions apply.

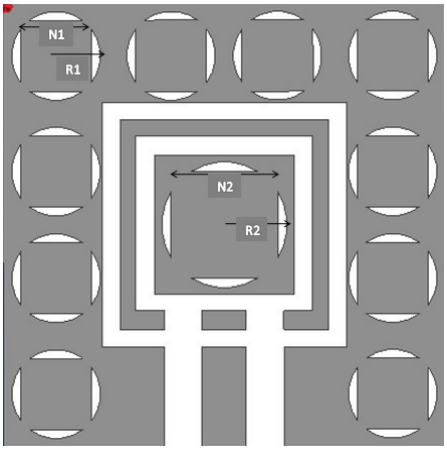


Fig.3 (a) Iteration 1

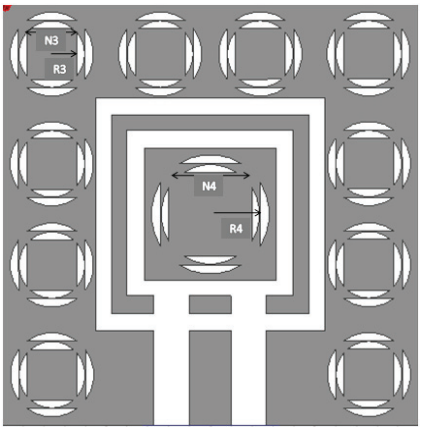


Fig.3 (b) Iteration 2

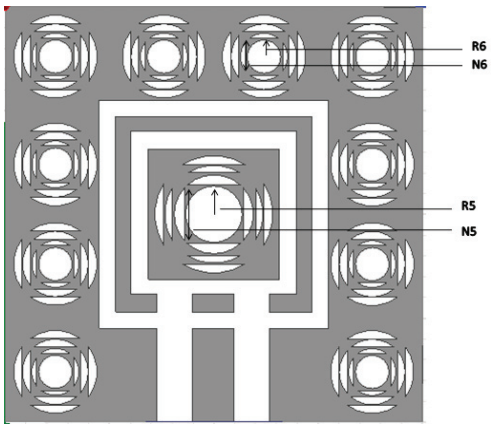


Fig.3 (c) Iteration 3

TABLE II. PARAMETRIC VALUES OF THE FRACTAL ANTENNA.

S. No.	Parameters	Values
1.	R1	1 mm
2.	N1	1.6 mm
3.	R2	1.4 mm
4.	N2	2.4 mm
5.	R3	0.8 mm
6.	N3	1.2 mm
7.	R4	1.2mm
8	N4	2 mm
9.	R5	0.6 mm
10.	N5	0.9 mm
11.	R6	1 mm
12.	N6	1.4 mm

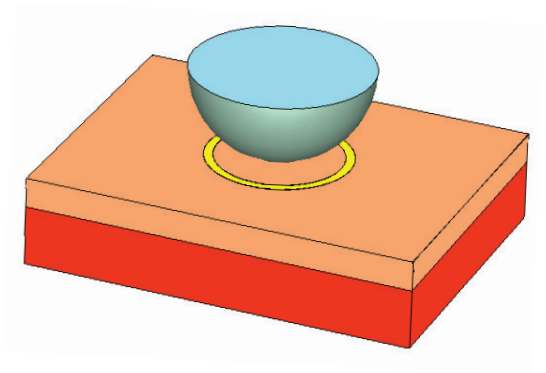


Fig.4. The proposed antenna with a hemispherical Lens and a parasitic ring over the body model.

V. DISCUSSION AND MEASUREMENT

A. Comparative analysis of the different iteration stages of fractal antenna.

The design is simulated through the Ansys HFSS 15.0 software, over a frequency range of 1 GHz to 5 GHz. The return loss plot, S11 parameter graph for all the three iterations combined is shown in figure.5(a). VSWR plot for all the iterations is shown in figure.5(b) From the values of the graph, the fractional bandwidth is calculated. Table III shows the comparison of the three iterations on the basis of gain, fraction bandwidth, return loss and SAR [4].

TABLE III. PERFORMANCE COMPARISON OF THE THREE ITERATIONS OF THE DESIGNED FRACTAL ANTENNA

Parameters	Iteration 1	Iteration 2	Iteration 3
Gain (dB)	-13.3	-13.3	-13.4
Fractional Bandwidth	27.45 %	60.46 %	69.76 %
Return loss (dB)	-32.6	-26.7	-28.05
Frequency Bands	(2.2-2.9) GHz ISM band	(1.5-2.8) GHz ISM band	(1.4-2.9) GHz WMTS, ISM band
SAR at power 1 Wattt	405.2	404.08	408.3

From the table we can infer that with the increase in the number of iterations, fractional bandwidth has significantly improved. The final antenna shows ultra-wideband characteristics that can work for two different medical bands namely ISM (Industrial, scientific and medical) band and WMTS (Wireless Medical Telemetry Service) band. Also, not much of a change is observed in gain, SAR and return loss values

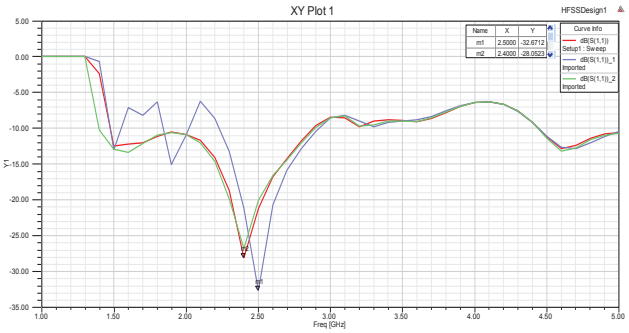


Fig.5 (a) Return loss plot for different iterations of the antenna.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:08:57 UTC from IEEE Xplore. Restrictions apply.

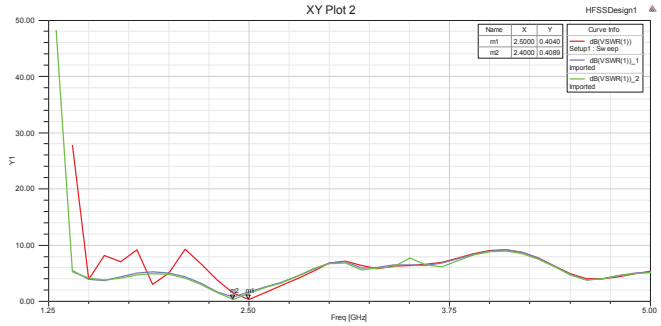


Fig.5 (b) VSWR plot for different iterations of the antenna.

B. *Gain enhancement of the implantable fractal antenna.*

1) *Encapsulation coating around the antenna.*

From table III, we can infer that the final iteration of fractal antenna was having gain -13.4 dB. To improve this, a thin layer of Teflon with dielectric constant 2.08 is coated around the antenna. After this ,an improvement in gain of 2.3 dB is obtained. So, the gain becomes -11.1 dB as shown in Figure.6 (b).

2) *External Parasitic ring.*

By coupling the exiting electromagnetic wave to an external parasitic ring, antenna performance can be improved [2]. This ring is placed on the skin with radius 24mm and thickness 3 mm. It has improved gain by 5.2 dB. So, the gain becomes -5.98 dB as shown in Figure.6 (c).

3) *External Hemispherical Lens*

To improve the implanted antenna performance an external hemispherical lens was designed and investigated to be placed on the skin over the implanted antenna. It was made up of pyrex material with dielectric constant 4.82. To minimize the spreading of incident divergent energy in undesired directions, Lens antennas are used [2]. They collimate incident energy-producing highly directive beams by converting divergent energy waves to plane waves. It has improved gain by 1.45 dB. So, the gain becomes -4.53 dB as shown in Figure.6 (d).

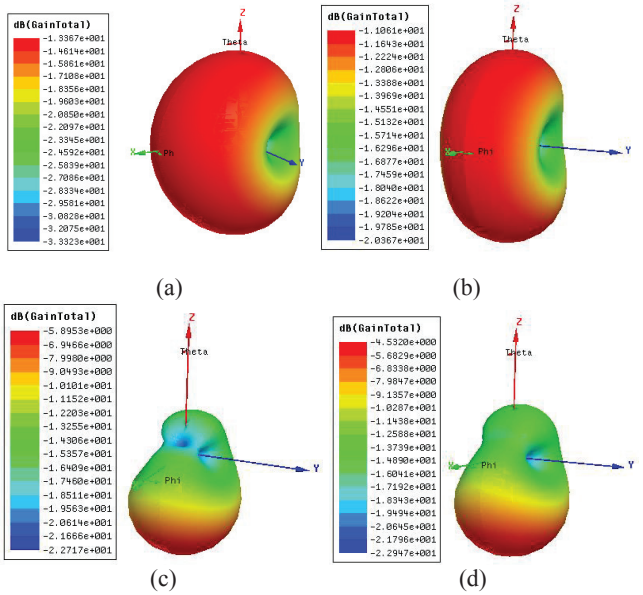


Fig.6 3-D polar plot showing gain of the antenna with various gain enhancement techniques used. (a) Third iteration fractal antenna (b) After encapsulation c) With external parasitic ring. (d) With hemispherical lens.

After using all the gain enhancement techniques together in the final iteration of the fractal antenna, a total of 8.8 dB of gain has been improved. Return loss plot, SAR and 2-D radiation patterns of the antenna has been shown in Figure.7,8 and 9 respectively.

Observing the return loss plot, return loss was found to be -14.4 dB at first resonating frequency 2.2 GHz and - 11.5 dB at second resonating frequency. Impedance bandwidth of the first band (3-2 GHz) is 1 GHz and of the second band (4.4-4.8 GHz) is 0.4 GHz. From these values, fractional bandwidth is calculated that is 37.13 % (wideband) for the first band and 8.7 % of the second band.

SAR (specific absorption rate) when the input power was 1 Watt is 401.8 W/kg. India has adopted the most strict international norms for wireless communication [8]. All the new designs of wireless instruments shall realize with the SAR values of 1.6 W/kg averaged over 1 gram of human tissue. On reducing the input power to 3.9 mW, the SAR value is 1.41 W/kg which comes under the minimum allowed SAR value. Far-field radiation pattern showing the E-plane and H-plane patterns at the operating frequency (2.45 GHz) and the resonating frequency (2.2 GHz).

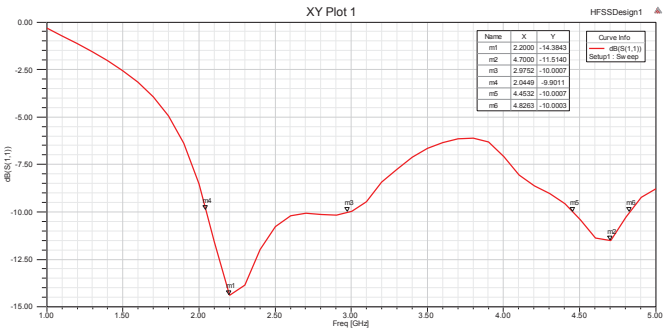


Fig.7 Return loss plot (S11 parameter)

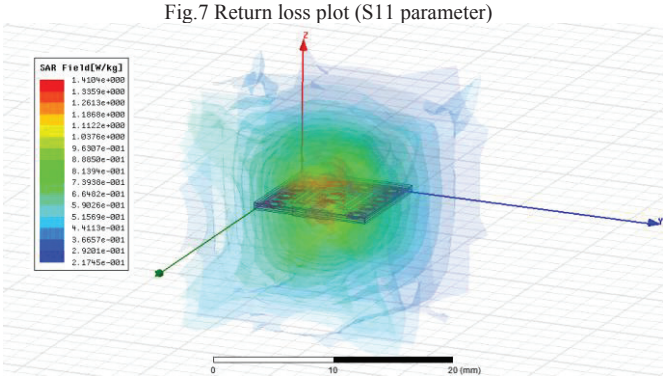


Fig.8 SAR (specific absorption rate)

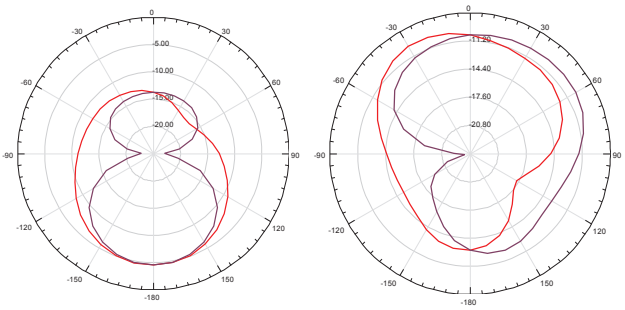
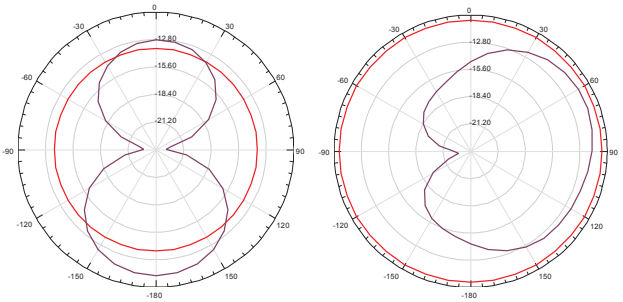


Fig.9 E-plane far field pattern: (a) at operating frequency 2.45 GHz (b) at resonating frequency 2.2 GHz

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:08:57 UTC from IEEE Xplore. Restrictions apply.



(c) (d)

Fig.9 H-plane far field pattern: (c) at operating frequency 2.45 GHz (d) at resonating frequency 2.2 GHz

VI. CONCLUSION

The design and analysis of a biomedical implantable antenna is addressed in this paper. With the help of the fractal geometry, the antenna is miniaturized and show wideband characteristics as per the requirement for the implantable antenna. The materials for substrate and superstrate are assigned keeping in mind the flexibility and biocompatibility of the antenna. Due to the lossy environment created by human tissue, the gain was low. So, various gain enhancement techniques are used namely encapsulation, external parasitic ring and lens. A gain enhancement of 8.8 dB has been observed combining all the three techniques. It has also been observed that due to the addition of these enhancement techniques, the SAR value of the antenna doesn't change significantly**.**

ACKNOWLEDGMENT

The author would like to thank the ``Research project grant to the faculty of DTU" under Delhi Technological University for its financial support of this work. The authors would also like to thank the anonymous reviewers for the feedback.

REFERENCES

[1] M. Ramya and R. Rani,"A compendious review onfractal antenna geometries in wireless communication" proceedings of the fifth international conference on initiative computation technologies (ICICT-2020) IEEE Xplore Part Number: CFP20F70-ART; ISBN:978-1-7281-4685-0.

[2] S. Das and D. Mitra, "A compact wideband flexible implantable slot antenna design with enhanced gain" IEEE Transactions on Antennas and Propagation (Volume: 66, Issue: 8, Aug. 2018) DOI-10.1109/TAP.2018.2836463, INSPEC Accession Number: 18006005.

[3] A. Tripathi, P. Rao and R. Mishra, “An AMC Inspired Wearable UWB Antenna for Skin Cancer Detection”, 2020 International Conference on Electrical and Electronics Engineering (ICE3-2020).

[4] A. Sharif, J. Ouyang, H. Chattha, M. Imran and Q. Abasi, “Wearable UHF RFID Tag Antenna Design using Hilbert Fractal Structure”, 978-1-7281-2797-2/19/$31.00 ©2019 IEEE.

[5] S. Alamri ,"Implanted antennas for Biomedical applications", University of Sheffield, Identification number uk.bl.ethos.701748.

[6] S. Suganthi, Dr. S. Raghavan, Dr. D. Kumar and Dr. J. Arputha Vijaya Selvi, " Study of Compact Hilbert Curve Fractal Antennas for Implantable Medical Applications ", International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, Volume 2.

[7] T. Das and S. Behera," Design of a Compact antenna for Biomedical Applications" IEEE International Symposium on Antennas and Propagation(APSYM2016), Cochin University of Science and Technology (CUSAT), Cochin, India, 15-17 December 2016.

[8] M. Telkapalli and K. Kavya," Optimization Techniques for Improving the Bandwidth of Biomedical Antennas", Journal of Green Engineering (JGE) Volume-10, Issue-10, October 2020.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:08:57 UTC from IEEE Xplore. Restrictions apply.

2022 2nd Asian Conference on Innovation in Technology (ASIANCON)
Pune, India. Aug 26-28, 2022

Blocking Probability Analysis for Energy Efficient Next Generation Green Cognitive Radio Networks

Akanksha Srivastava*
ECE Department
Delhi Technological University
Delhi, India
akankshasrivastava30@gmail.com

Gurjit Kaur
ECE Department
Delhi Technological University
Delhi, India
gurjitkaur@dtu.ac.in

2022 2nd Asian Conference on Innovation in Technology (ASIANCON) | 978-1-6654-6851-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ASIANCON55314.2022.9908767

***Abstract*—In the present scenario, Cognitive Radio Networks (CRNs) plays an important role in designing the 5G and B5G wireless communication networks. In next-generation networks, there will be a large of users, and this huge data requirement demands improvement in the quality of service (QoS) and energy efficiency (EE) and decrements in blocking probability. The CRNs technology help to solve the major issues of $5G$, like seamless connectivity, call blocking and power management, etc. In this work, we analyze the blocking probability, outage probability, and energy efficiency of cognitive users (CUs) in the CRNs. These parameters are envisioned as key parameters to examine the overall network performance. We analyze the blocking probability of a network where a user is blocked due to the unavailability of resources. The simulation results reveal the performance of CUs in terms of EE and BP. The analysis presented in this work will be helpful for the deployment of energy-efficient ultra-dense network because results show that the CU's EE improves with the increase in availability of resources of primary networks.**

***Index Terms*—Cognitive radio networks, Green communication, Energy-efficiency, Blocking probability, Outage probability.**

I. Introduction

Due, to the explosive expansion of mobile users, energy consumption in the field of wireless communication has been raised dramatically in recent years. Information and communication technologies (ICTs) are accountable for 2% of greenhouse gas emissions and 3% to 6% of overall global energy consumption [1]. In order to achieve green communications, wireless systems must fulfill data rate requirements with low energy consumption. According to the Federal Communications Commission (FCC), almost 60% of the spectrum band is unused in the dedicated spectrum bands [2]. Cognitive Radio Networks (CRNs) are an emerging technology in which users of one network can utilize the unused resources of another network in an opportunistic manner [3]. The increased demand for real-time and multimedia services engage the network's channel for a prolonged time, which leads the base stations (BSs) to reject new incoming calls, and it causes an increase in blocking probability (BP) [4]. The BP increases when all the network resources are utilized, while outage probability (OP) appears because of the carrier-to-interference ratio and carrier-to-noise ratio. This situation degrades the network performance, and it is caused by inappropriate spectrum management, improper traffic load balancing, and inefficient resource utilization [5].

In the existing work, various authors suggested different methods to improve the system performance, mainly focused on enhancing data rate and throughput but in this work a joint analysis of BP, outage probability (OP), and energy efficiency (EE) have been presented. The mobile communication system should follow an efficient resource utilization scheme to minimize BP. For solving the spectrum/channel utilization problem, the CRNs technology is the best option that permits the cognitive users (CUs) (users of secondary network or unlicensed band) to access the unused licensed spectrum bands of the primary network (PN). The performance of CUs depends on the availability of PN's resources and the traffic of PUs. In this work, we analyze the performance of CUs in terms of BP, OP, and EE in the presence of various parameters of PN.

II. Related Works

Researchers and academicians have done prominent work for promoting energy-efficient CRNs in several aspects. In [6], the authors introduced a Markov chain-based spectrum access scheme and investigated BP. An intelligent resource allocation mechanism has been discussed in [7] to achieve the reduction in BP, but the energy-saving perspective is not considered. In [8] and [9], the authors presented a scheme for backhaul load balancing in next generation heterogeneous networks. The simulation results of this work reveal that the proposed scheme performs in terms of BP and QoS. In [10], a multi-objective resource allocation problem is developed to optimize the EE in heterogeneous CRNs. In [11], the authors presented a dynamic radio resource allocation (DRRA) mechanism that helps to estimate the BP and EE of the CRNs. Authors in [12] have been presented a Q-Learning based algorithm for analyzing the BP of the network. A Learning automata-based mechanism has been discussed in [13], which calculates the BP and spectral efficiency of CUs in CRNs. Although some of these literary works, the energy-saving issue has been studied, but our analysis is different in various aspects. Specially, we examine the three important parameters; BP, OP and EE in the context of CUs.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:16:25 UTC from IEEE Xplore. Restrictions apply.

A. Contributions

The key objective of this work is to collectively estimate the influence of energy-saving on BP and OP for the scenario of CRNs. The significant contributions of this work outlined as follows:

- Design a system model and analyze the blocking probability of CU that is blocked due to the availability of PN's resource. The BP of CUs is estimated in terms of data traffic intensity and number of channels.
- The outage probability of system is investigated in terms of number of channels at a variable data traffic of PU.
- The energy efficiency of CUs is analyzed in terms of number of channels at a variable data traffic of PUs and traffic arrival rate of CUs.

B. Paper Organization

The paper's progression is as follows: Section II presents the existing related works. In Sections III, the proposed network system model is with presented. The performance evaluations and results are discussed in Section IV. Finally, the work is concluded in Section V.

III. System Model

A. Analysis of blocking probability

Definition-*The blocking probability arises when all the resources of network (defined numbers during the network deployment) are already engaged by other users.*

For a network, least number of channels needed per cell is calculated using Erlang-B [14]. Erlang-B estimates probability under M/M/n/n, having some assumptions as follows. (i) If all resources are engaged, then new arriving user's request will be blocked and blocked request discarded (no queue formation), (ii) Users' request arrival follow the Poisson process, and the number of available resources (i.e., n) in the network is known.

The M/M/n/n model is presented in Fig. 1, where there is a PN and it is assumed that n is equal to 4 (we can not show large n value due to space limitation). Fig. 1 shows that the first 4 CUs obtain resources (channels). But, because of scarcity of resources, the fifth user has been blocked. In this context n and $g(t)$ represent the number of available and occupied channels respectively, at the time t. Hence, if $g(t) = n$ then the request of the new user will be discarded and if $g(t) < n$, then the request will be accepted. The Erlang's function is utilized to estimate the BP in an individual channel that has $\mathbb{C}_n$ capacity and offered network traffic data with intensity D. This approach is applied for the calculation of probability of a user's call request that is blocked by the network because of the channel's unavailability. Here, for calculating the BP over various channels which have various data traffic streams in the network, the following parameters are used. (see Table I for a summary of notations used)

For low traffic the BP can be determined as $BP_x = \sum_{n \in Q} BP_n$ [15]. But, if channel blocking is independent, then BP can be determined as:

$$1 - BP_x = \prod_{n \in Q} (1 - BP_n) \tag{1}$$

TABLE I: Summary of Notation

Notation	Description
n	Channels, with $\mathbb{C}_n$ (trunks)
x	Data traffic stream
Q_x	Route of data traffic stream x, where $n \in Q$; when n is on the route of x
BP_n	Blocking probability in channel n
BP_x	Blocking probability of data stream x
D_x	Offered data traffic intensity of x
D_n	Offered data traffic intensity to channel n

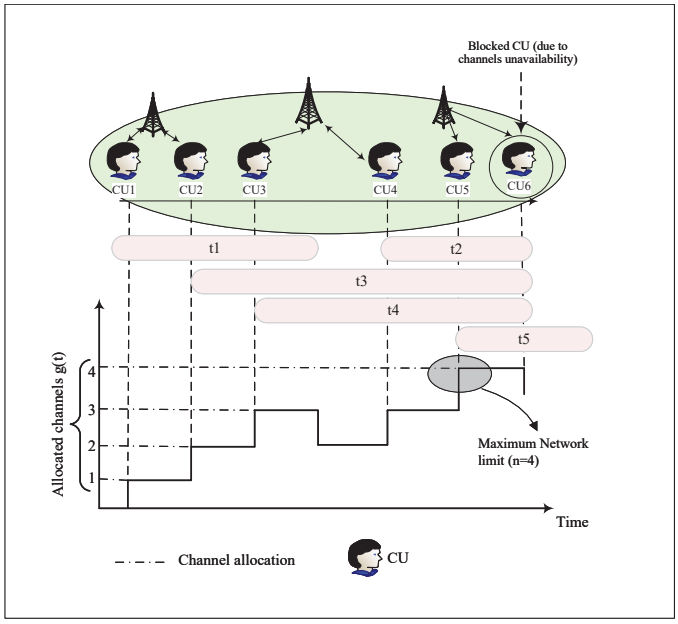


Fig. 1: System model

After this, Erlang's fixed-point process is used to evaluate the BP_n. In the case of low blocking, the offered data traffic intensity to channel n can be estimated by Erlang formula:

$$D_n = \underbrace{\sum_{x:n \in Q_n} D_x}_{\substack{sum\ over\ the\ data \\ streams\ using\ channel\ n}}, \qquad BP_n = E_f(\mathbb{C}_n, D_n) \tag{2}$$

Now, for the case of high number of blocking, the considered traffic thinning in available channels is:

$$\begin{cases} D_x^n & thinned\ traffic\ provided\ by\ stream\ x\ to\ channel\ n \\ D_n = \sum\limits_{x:n \in Q_n} D_x^n & addition\ of\ thinned\ traffic\ provided\ to\ channel\ n \end{cases} \tag{3}$$

$$D_x^n = \begin{cases} 0, & n \notin Q_n \\ D_x \cdot \underbrace{\prod\limits_{g \in Q_x - \{n\}} (1 - BP_g)}_{\substack{In\ other\ channel\ thinning \\ factor}}, & n \in Q_n \\ = D_x \frac{1}{1 - BP_n} \prod\limits_{g \in Q_n} (1 - BP_g) & \end{cases} \tag{4}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:16:25 UTC from IEEE Xplore. Restrictions apply.

Therefore, for n channels the BP can be written as

$$BP_n = E_f\left\{\mathbb{C}_n, \sum_{x:n\in Q_n} D_x \frac{1}{1-BP_n}\prod_{g\in Q_x}\left(1-BP_g\right)\right\} \quad (5)$$

The blocking related to channel n and other available channels are associated with each other. For analyzing the BP, initially, we consider $BP_n = 0, \forall n$ and substitute this in (5) on right hand side (RHS). Then we get a new value of BP_n, again substitute this value into the RHS and continue this process for all channels. Erlang's fixed-point process refers for solving the set of equations. By putting the value of BP_n on (6)

$$BP_x = 1 - \prod_{n\in Q_n}\left(1-BP_n\right) \quad (6)$$

The BP of a CUs on the traffic intensity of PU and number of resources. If the number of resources increase then the BP decreases. Therefore, if the available resources are properly utilized than the BP of CUs can be minimized.

B. Analysis of Outage probability

In this work, Rician fading and Nakagami fading [16] are considered to estimate the outage probability. It is assumed that F_p represents immediate faded power of the required signal, and $\rho_p = E_{F_p}\left[F_p\right]$ indicates their short-term average.

In this context, F_p depends on either Nakagami (having Nakagami factor $\Im_p$) or Rician (having Rician parameter $\wp_p$) kind of fading based on the situation of the moment. Now in the case of Nakagami fading the pdf of F_p is written as:

$$p_{F_P}\left(F_p\right) = \left(\frac{\Im_p}{\rho_p}\right)^{\Im_p}\frac{\left(F_p\right)^{\Im_p-1}}{\Gamma\left(\Im_p\right)}\exp\left(-\frac{\Im_p F_p}{\rho_p}\right) \quad (7)$$

In the case of Rician fading, the pdf is shown by (8),

Where $I_m(\cdot)$ is the m^{th} order-Bessel function. Notably, $\Im_p = 1$ and $\wp_p = 0$, directed towards Rayleigh case having pdf $p_{F_P}\left(F_p\right) = \exp\left(-F_p/\rho_p\right)/\rho_p$. Note that the power $U_{\mathbb{Z}}$ of operating interference signals $\overset{U_{\mathbb{Z}}}{_{z=1}}\{F_z\}$ is considered to be equal to $\rho_{\mathbb{Z}}$. The overall instantaneous power $F_{\mathbb{Z}}$ is the addition of the instantaneous powers of the $U_{\mathbb{Z}}$, which is measured as $F_{\mathbb{Z}} = \sum_{z=1}^{U_{\mathbb{Z}}} F_z$.

In the system the outage probability with interference and limited power can be written as carrier-to-interference ratio (CIR); $\lambda = \frac{F_P}{F_{\mathbb{Z}}}$, which lies below a predefined protection ratio λ_{th}. It can be considered as instantaneous power of wanted signals F_P, which is not more than the predefined threshold F_{th}. Mathematically, the OP can be determined as,

$$OP_{ut} = P\left[\lambda = \frac{F_P}{F_{\mathbb{Z}}} \le \lambda_{th} or F_P \le F_{th}\right] \quad (9)$$

The (9) can be rewritten in terms of F_P, $F_{\mathbb{Z}}$, $p_{F_{\mathbb{Z}}}\left(F_{\mathbb{Z}}\right)$ is,

$$OP_{ut} = 1 - \int_{F_{th}}^{\infty}\left[\int_0^{F_p/\lambda_{th}} p_{F_{\mathbb{Z}}}\left(F_{\mathbb{Z}}\right)dF_{\mathbb{Z}}\right]p_{F_P}\left(F_p\right)d_{F_P} \quad (10)$$

Assumed that signals received from users are Rician distributed with ρ_p and $\wp_p$, and $U_{\mathbb{Z}}$ interference with average fading power $\rho_{\mathbb{Z}}$ in the system. Therefore, the overall instantaneous power $[F_{\mathbb{Z}}]$ have pdf as:

$$p_{F_{\mathbb{Z}}}\left(F_{\mathbb{Z}}\right) = \frac{{F_{\mathbb{Z}}}^{U_{\mathbb{Z}}-1}}{{\rho_{\mathbb{Z}}}^{U_{\mathbb{Z}}}\left(U_{\mathbb{Z}}-1\right)!}\exp\left(-\frac{F_{\mathbb{Z}}}{\rho_{\mathbb{Z}}}\right) \quad (11)$$

Implementing the gamma cumulative distribution, and consider (8) into (10). It provides the resulting OP with shifting the order of integration and summation, thus the OP can be indicated by (12);

In (12), the term $Q_1(\cdot,\cdot)$ represents the Marcum Q- function. Now applying a relevant updation then the (12) can be written as:

where K_i can be estimated as:

$$H = \sqrt{2\wp_p\left(1+\frac{\rho_p}{\left(\wp_p+1\right)\lambda_{th}\rho_{\mathbb{Z}}}\right)^{-1}} \quad (15)$$

$$J = \sqrt{2\left(1+\rho_p+\frac{\rho_p}{\lambda_{th}\rho_{\mathbb{Z}}}\right)\frac{F_{th}}{\rho_p}} \quad (16)$$

$Q_{X,i}(\cdot,\cdot)$ is the Nuttall Q-function.

$$Q_{X,i}(H,J) = \int_J^{\infty} W^X \exp\left(-\frac{W^2+H^2}{2}\right)I_i(HW)dW \quad (17)$$

$$Q_X(H,J) = \int_J^{\infty}\left(\frac{W}{H}\right)^{X-1} * \exp\left(-\frac{H^2+W^2}{2}\right)I_{X-1}(HW)\,WdW \quad (18)$$

$$= Q_1(H,J) + \exp\left(-\frac{H^2+W^2}{2}\right) * \sum_{\Omega}^{X-1}\left(\frac{J}{H}\right)I_{\Omega}(H,J) \quad (19)$$

IV. PERFORMANCE EVALUATION

In this section, we present the simulation results to illustrate the various parameters like BP, OP, and EE for CRNs. MATLAB software is used for the simulation purpose. It is assumed that there are 25 users in the network, and these users are distributed randomly, and generate the traffic using Poisson process for estimating the BP. (see Table II for other considered parameters)

TABLE II: Simulation Parameters

Name of Parameter	Parameters value
Transmission power	30 W
Operating frequency	2 GHz
Data traffic rate	20 – 40 $Mbps$
Antenna gain	15 dB
User speed	30 Km/h
System bandwidth	15 MHz
No of PUs	20

Fig. 2 presents that the BP of the network is changed due to the variation in the offered traffic load. The offered traffic (D_n) is in Erlang, and n is the number of channels. From Fig. 2, it is noted that the data traffic system gives non-linear characteristics and with an increase in the number of channels (resources) n, the BP will decrease for the given value of the offered traffic load.

Fig. 3 indicates the investigation of outage probability at different data traffic of PUs for various number of channels.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:16:25 UTC from IEEE Xplore. Restrictions apply.

$$p_{F_P}\left(F_p\right) = \frac{\wp_p + 1}{\rho_p} \exp\left(-\wp_p - \frac{\wp_p + 1}{\rho_p} F_p\right) * I_m\left(2\sqrt{\frac{\wp_p\left(\wp_p + 1\right)}{\rho_p} F_p}\right) \tag{8}$$

$$\begin{aligned} OP_{ut} = 1 - Q_1\left(\sqrt{2\wp_p}, \sqrt{\frac{2\left(\wp_p + 1\right) F_{th}}{\rho_p}}\right) + \frac{\wp_p + 1}{\rho_p} \sum_{i=0}^{U_{\mathbb{Z}}} \frac{1}{i} * \int_{F_{th}}^{\infty} \left(\frac{F_p}{\lambda_{th}\rho_{\mathbb{Z}}}\right)^i \\ * \exp\left[-\wp_p - \left(\frac{1}{\lambda_{th}\rho_Z} + \frac{1 + \wp_p}{\rho_p}\right) F_p\right] * I_m\left(2\sqrt{\frac{\wp_p\left(\wp_p + 1\right) F_p}{\rho_p}}\right) d_{F_P} \end{aligned} \tag{12}$$

$$OP_{ut} = 1 - Q_1\left(\sqrt{2\wp_p}, \sqrt{\frac{2\left(\wp_p + 1\right) F_{th}}{\rho_p}}\right) + \frac{H^2}{2\wp_p} \sum_{i=0}^{U_{\mathbb{Z}}-1} \frac{K_i}{i!} Q_{2i+1,0}\left(H, J\right) \tag{13}$$

$$K_i = \left(2 + \frac{2\left(\wp_p + 1\right)\lambda_{th}\rho_{\mathbb{Z}}}{\rho_p}\right)^{-i} \times \exp\left[-\wp_p + \wp_p\left(1 + \frac{\rho_p}{\left(\wp_p + 1\right)\lambda_{th}\rho_{\mathbb{Z}}}\right)^{-1}\right] \tag{14}$$

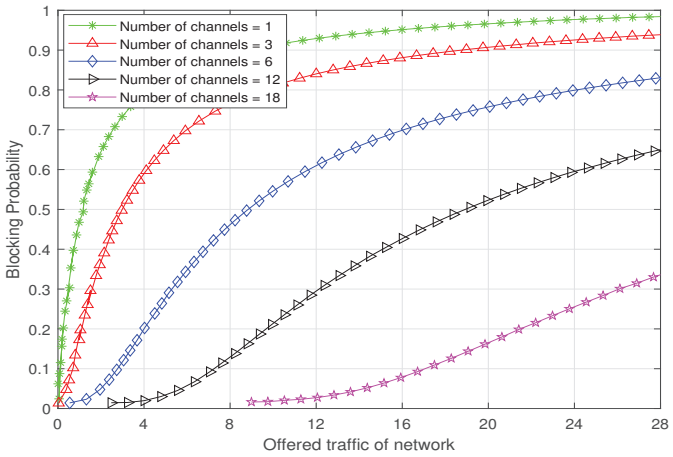


Fig. 2: Blocking probability with respect to offered traffic for multiple number of channels.

The outage probability decreases with an increase in the number of channels in the network. The reason behind this when the number of channels are increased, then the CU has more opportunities to select the Channel. In line with, the outage probability is also higher for low traffic of rate of PUs.

The impact of the PN's density (the number of channels) on the EE of the CU is examined in Fig. 4. From Fig. 4, it can be observed that the CU gets higher EE if there is more number of channels in the PN because CU has more options to pick a channel with a smaller traffic load or directly select a vacant channel. Fig. 4 indicates that with an increase in the number of channels in the PN, the EE of CU is also increases for a low value of PU's traffic load. Because of low PU's traffic loads, the CU spends a low amount of energy for accessing the channel and can detect an appropriate channel after a few time slots. It results in a larger EE for the CU.

The simulation result reveals that with an increase in the CU's packet arrival rate, the EE of CU degrades, as shown in Fig. 5. From Fig. 5, it can be observed that the EE of CU is increased by increasing the channel's density in PN, but system performance is diminished by increasing packet arrival rate of CU. The explanation is as follows when the packet arrival rate is high; then the CU has to send more packets, which produces a larger time for the data transmission. It decreases the opportunity to choose an appropriate channel for data transmission and therefore extends its energy expenditure. Hence, we can conclude that the increment in the packet arrival rate of CU puts negative impacts on EE of CU.

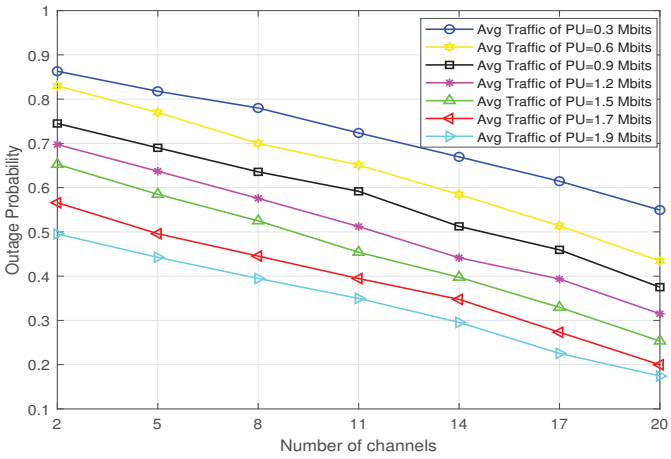


Fig. 3: Outage Probability with respect to the number of channels.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:16:25 UTC from IEEE Xplore. Restrictions apply.

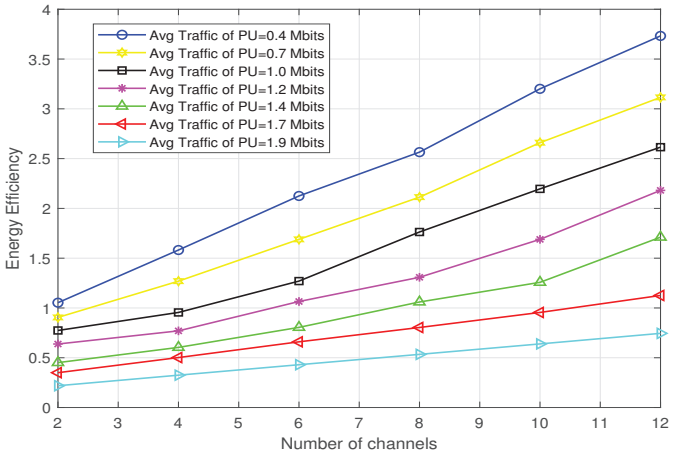


Fig. 4: Energy efficiency of CU for various average traffic load of PUs.

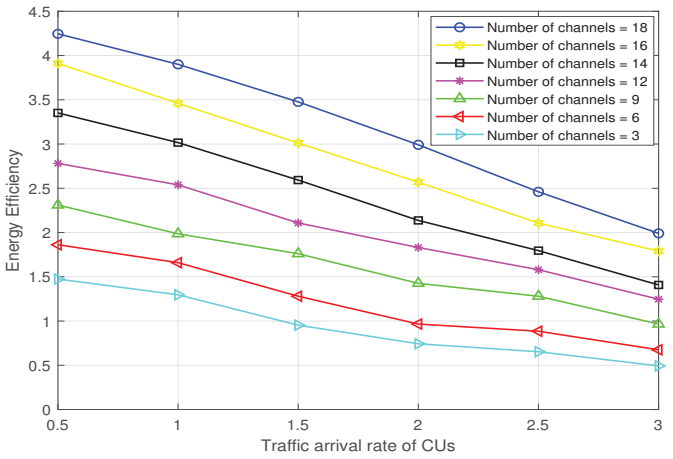


Fig. 5: Energy efficiency of CU for various number of channels.

V. Conclusion

The next-generation network will contain more mobile users, which results in higher interference and extended resource allocation. It leads toward more blocking probability and significantly affects the system performance and energy efficiency. This work investigates the outage probability and blocking probability for CRNs and analyses the impact of these parameters on the energy efficiency of CU. Our work has contributed significant insight in terms of blocking probability, energy efficiency, and outage probability, and it motivates us to design the "Future Green Communications System".

Acknowledgment

This work is supported by Women Scientist Scheme-A, Department of Science and Technology, New Delhi, Government of India under Grant File No: SR/WOS-A/ET-154/2017. The authors would like to thank the reviewers for their feedback.

References

[1] K. Imasiku, V. Thomas, and E. Ntagwirumugara, "Unraveling green information technology systems as a global greenhouse gas emission game-changer," *Administrative Sciences*, vol. 9, no. 2, p. 43, 2019.

[2] A. Srivastava, M. S. Gupta, and G. Kaur, "Energy efficient transmission trends towards future green cognitive radio networks (5g): Progress, taxonomy and open challenges," *Journal of Network and Computer Applications*, p. 102760, 2020.

[3] M. S. Gupta and K. Kumar, "Afsos: An auction framework and stackelberg game oriented optimal network's resource selection technique in cognitive radio networks," *IEEE Transactions on Network and Service Management*, 2021.

[4] A. Srivastava and G. Kaur, "Resource management for traffic imbalance problem in green cognitive radio networks," *Physical Communication*, vol. 48, p. 101437, 2021.

[5] A. Srivastava, M. S. Gupta, and G. Kaur, "A game theory based approach for opportunistic channel access in green cognitive radio networks," in *2019 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*. IEEE, 2019, pp. 1–6.

[6] X. Ge, B. Yang, J. Ye, G. Mao, C.-X. Wang, and T. Han, "Spatial spectrum and energy efficiency of random cellular networks," *IEEE Transactions on Communications*, vol. 63, no. 3, pp. 1019–1030, 2015.

[7] M. S. Gupta and K. Kumar, "Application aware networks' resource selection decision making technique using group mobility in vehicular cognitive radio networks," *Vehicular Communications*, vol. 26, p. 100263, 2020.

[8] Y. L. Lee, T. C. Chuah, A. A. El-Saleh, and J. Loo, "User association for backhaul load balancing with quality of service provisioning for heterogeneous networks," *IEEE Communications Letters*, vol. 22, no. 11, pp. 2338–2341, 2018.

[9] M. S. Gupta, A. Srivastava, and K. Kumar, "Seamless vertical handover for efficient mobility management in cooperative heterogeneous networks," in *Proceedings of the 2nd International Conference on Data Engineering and Communication Technology*. Springer, 2019, pp. 145–153.

[10] S. Wang and C. Wang, "Joint optimization of spectrum and energy efficiency in cognitive radio networks," *Digital Communications and Networks*, vol. 1, no. 3, pp. 161–170, 2015.

[11] S. Batabyal and S. S. Das, "Distance dependent call blocking probability, and area erlang efficiency of cellular networks," in *2012 IEEE 75th Vehicular Technology Conference (VTC Spring)*. IEEE, 2012, pp. 1–5.

[12] J. Chen, Y. Wang, Y. Li, and E. Wang, "Qoe-aware intelligent vertical handoff scheme over heterogeneous wireless access networks," *IEEE Access*, vol. 6, pp. 38 285–38 293, 2018.

[13] G. A. Beletsioti, G. I. Papadimitriou, P. Nicopolitidis, E. Varvarigos, and S. Mavridopoulos, "A learning-automata congestion-aware scheme for energy-efficient elastic optical networks," *IEEE Access*, 2020.

[14] V. B. Iversen *et al.*, "Teletraffic engineering handbook," *ITU-D SG*, vol. 2, p. 16, 2005.

[15] M. Gupta and K. Kumar, "Progression on spectrum sensing for cognitive radio networks: A survey, classification, challenges and future research issues," *Journal of Network and Computer Applications*, vol. 143, pp. 47–76, 2019.

[16] M. S. Gupta and K. Kumar, "Group mobility assisted network selection framework in 5g vehicular cognitive radio networks," *Physical Communication*, vol. 51, p. 101578, 2022.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:16:25 UTC from IEEE Xplore. Restrictions apply.

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1746-8809.htm

Board committees and financial performance: exploring the effects of gender diversity in the emerging economy of India

Shubham Singhania, Jagvinder Singh and Deepti Aggrawal
University School of Management and Entrepreneurship, Delhi Technological University, Delhi, India

Received 23 March 2022
Revised 9 July 2022
30 August 2022
Accepted 16 September 2022

Abstract

Purpose – This study aims to highlight the impact of introducing women directors to board committees, thereby empowering them to contribute to decision-making, and as a result, influence firms' financial performance in an emerging economy.

Design/methodology/approach – This study uses a fixed-effects panel data regression model to test the impact of gender diversity on corporate boards as well as board committees on firms' financial performance. Two widely used diversity measures, the Blau index and the Shannon index, have been used to enhance the robustness of the results.

Findings – The findings suggest that gender diversity on prominent board committees (remuneration committee and nomination committee) positively affects firms' financial performance when measured by the market-based performance measure, but it is insignificant when measured through accounting-based performance indicator. Furthermore, the benefits of gender diversity accrue to the firms only when women are part of prominent committees and are engaged in governance mechanisms, rather than just being appointed on corporate boards as a means of tokenism.

Originality/value – This study is among the first to investigate the relationship between gender diversity and financial performance through the lens of committee assignments. Moreover, the unique cultural and institutional setting offered by India, which is an emerging economy, provides a fertile ground for understanding the role of women leaders in the workforce.

Keywords Women directors, Board committee, Women on board, Gender diversity, Financial performance, Panel data regression, Tobin's *Q*, India

Paper type Research paper

1. Introduction

The issue of gender diversity on corporate boards has received substantial momentum in recent years through public debates, news reports, global discussion forums as well as academic research (Chen and Kao, 2020; Ararat and Yurtoglu, 2021). The growing legal and mandatory provisions concerning the appointment of women directors on boards have further attracted the attention of the masses toward this domain. Despite the growing consciousness and directives adopted in various countries for enhancing gender diversity on corporate boards, the global representation of women is averaged at 24% (Credit Suisse, 2021).

There is an ongoing debate regarding the business case for women appointments on boards (Ain *et al.*, 2021; Ararat and Yurtoglu, 2021) because the research outcomes still suffer from various lacunas. For instance, on one hand, according to the Delhi High Court, there is an alarming increase in the number of false cases being reported by women (*Singh*, 2021), and on the other hand, women face various kinds of societal biasedness (Feng *et al.*, 2017; Jha and Jha, 2022) such as gender discrimination at the workplace, as well as harassment, which is clear with the growing complaints on the "Me Too" movement. To ensure that these lacunas are reduced, and companies can exploit the benefits derived from diversity, proper



International Journal of Emerging Markets
© Emerald Publishing Limited
1746-8809
DOI 10.1108/IJOEM-03-2022-0491

implementation of the rules and regulations must be ensured after thorough examination of the case of women appointments on board.

Moreover, attempts to establish a direct nexus between gender diversity and financial performance overlook the role of governance mechanisms in companies. Adams and Ferreira (2009), through their seminal work, emphasize that the presence of women directors on boards affects the governance of companies, which in turn shapes the financial performance. Multiple aspects of governance, such as attendance in board meetings, chief executive officer (CEO) compensation, and CEO turnover influence the extent to which gender diversity contributes to the commercial outcomes of a firm (Adams and Ferreira, 2009). The board of a company relies on the committees for its duties and assignments (Adams *et al.*, 2010). As a result, the individuals on these committees are responsible for steering the committee's decisions (Green and Homroy, 2018), while voicing their opinions. The active participation of women directors on board committees, therefore, can be more effective in monitoring the performance of the firms, than the mere presence of women directors on corporate boards. Moreover, the agency theory and the resource dependency theory of corporate governance provide justification for gender diversity on board for improvement in the decision-making, which shall have a bearing on financial performance.

Based on this backdrop, the objective of our study is to determine the relationship between gender diversity and financial performance by studying the moderating impact of committee assignments of women directors. The study has been conducted in the Indian context. This study makes multiple contributions to the literature. First, it is amongst the first few studies to investigate the relationship between gender diversity and financial performance through the lens of committee assignments. Though some studies have made the effort to determine the linkage between the board committees and financial performance (Berezinets *et al.*, 2017; Ames *et al.*, 2018), the role of gender diversity has not been examined in length. Second, the study shall provide new empirical evidence in the institutional setting of an emerging economy, India and allow comparison with similar studies in developed economies of the United States of America, the United Kingdom and Europe (Carter *et al.*, 2010; Green and Homroy, 2018). Based on the different institutional contexts, the results of various corporate governance aspects, especially gender diversity, are expected to vary across economies (Issa and Fang, 2019). Third, this study, in a culturally diverse and secular country like India, can provide a novel insight into how it responds to women's progression and gender representation on committees, and ultimately, its impact on the firm performance. This is because societal recognition and acceptability tend to depend on the status and position of an individual, rather than their professional achievements, and this recognition impacts the individual's ability to voice their opinions (Hubner *et al.*, 2022). Societal values and cultural aspects also play a prominent role in influencing the decision-making process as well as the perspectives of an individual (Martinsons and Davison, 2007). Hence, the governance of firms through committee appointments shall be impacted by the cultural aspects of a country.

This paper tries to address a relevant research gap by measuring the gender diversity on the remuneration committee as well as the nomination committee and further linking it with the financial performance. The rationale for choosing these committees is that apart from being constituted consistently across all firms as part of the corporate governance norms, they are essential in carrying out the major functions of the board. While other committees such as the audit committee are also essential, this paper does not consider them since a recent study in a similar setting found no significant impact of gender diverse audit committees on firm performance (Maji and Saha, 2021).

1.1 What differentiates India from other economies?

According to the World Economic Outlook Report 2021 (International Monetary Fund, 2021), India shall retain its tag of being the fastest-growing emerging economy. Following the global trend of legislations mandating a quota for women on board, India came up with a similar mandate in the year 2013. Furthermore, India's institutional and cultural environment, along with its greater informational transparency (Gupta *et al.*, 2019), provides a fertile ground to study gender diversity in workplaces.

Multiple factors differentiate India and make this emerging economy an interesting case for this study. First, India is one of the most culturally diverse countries in the world where the social structures are deeply rooted in the belief system (Schwartz, 2014). These social structures and cultural norms identify women with feminine roles and restrict them into family spheres. The high masculinity index of India is further evidence of the growing cultural stratification and strict division of roles among men and women (Jayachandran, 2021). The Indian social environment is dominated by the element of "hierarchical structure" (Gupta, 2017) which negatively impacts the freedom of voicing one's opinion in front of the senior executives, especially by the female employees who are already facing the clutches of femininity and patriarchal spill over. The hierarchical structure is further apparent by the fact that India scores higher than the global average in terms of the power distance index, which stands at 77 against the world average of 59 (Tripathi and Vijayan, 2020). Owing to self-promotion and self-enhancement in the high-power distance culture of India, women face a tough time gaining visibility in strategic positions (Gupta, 2017).

Second, gender discrimination is entrenched in the Indian system, and this can be attributed to the parent's gender-based stereotypical beliefs and expectations which impact the roles, attitudes and feelings of genders in the long run (Eckes and Trautner, 2012). This also has a bearing on the emotions that are built since childhood, such as fear, sadness and other powerless emotions in females and anger, pride, contempt, etc. in males (Abi Rached *et al.*, 2021). Women directors are often considered fit for doing routine tasks in the organizations and are not offered roles in prominent committees as compared to their male counterparts (Rebérioux and Roudaut, 2017). In the presence of this gender discrimination that restricts the independence of women in India (Kumar, 2020), women directors on boards would often feel reluctant to ask for roles on prominent committees. Since female directors are brought up in a culture where they are expected to be of "humane orientation", with feelings such as kindness, generosity, etc. (House *et al.*, 2004), they tend to show a greater inclination toward community service projects vis-à-vis the male directors (Groysberg and Bell, 2013). Hence, often, they are appointed to corporate social responsibility (CSR) committees, where they are believed to do well.

Third, the state of women's representation on Indian corporate boards is in a dismal position. India ranks 140 out of 156 countries in the global gender gap index, which indicates a high level of gender inequality (World Economic Forum, 2021). According to the report by Credit Suisse (2021), even after the constitution of mandatory provisions through the Companies Act 2013, the female representation on boards of large Indian companies has increased from 11.4% in 2015 to merely 17.3% in 2021, which is way below the global average.

The above factors lead to variations in the governance environments as well as the corporate governance practices of this emerging economy (Lattemann, 2014), thereby necessitating the study of the relatively unique case of India.

2. Review of the literature and hypothesis development

2.1 Theoretical background

The study uses the agency theory (Jensen and Meckling, 1976) and the resource dependency theory (Pfeffer and Salancik, 1978) to explain the nature of the relationship between gender

diversity and financial performance. The agency theory emphasizes that the fundamental function of a board is to manage the affairs of a company, which is embedded in the agency aspect (Carter *et al.*, 2010; Shahroor and Ismail, 2022), and a diverse board is likely to improve this function. Board diversity is found to enhance board independence (Carter *et al.*, 2003), which shall improve the level of decisions and thus firm value (Sarkar and Selarka, 2021). Hence, the agency theory highlights a beneficial aspect of board diversity.

Resource dependency theory acts as a connecting link between gender diversity and financial performance. The theory advocates that the firm accrues benefits from boards by making use of their advice, counseling, legitimacy and access to resources. The diversity in boards (gender, in this case) would reduce the dependency of decision-making on a few members and would increase the bundle of skills brought forth by women members in the form of values, leadership styles and linkages, which shall improve the decision-making process and reduce the dependence on the external environment (Nicolò *et al.*, 2021; Terjesen *et al.*, 2016).

2.2 Board gender diversity and financial performance

Ample studies have been undertaken to examine the impact of board diversity on firm outcomes (Baker *et al.*, 2020; Singh *et al.*, 2021b) and researchers at the global level have tried to capture the role played by women leaders in influencing firm performance (Duppati *et al.*, 2020; Martinez-Jimenez *et al.*, 2020; Brahma *et al.*, 2021). However, the results are mixed. One set of advocates have found a positive relationship between women's representation on board and financial performance (Bennouri *et al.*, 2018; Brahma *et al.*, 2021; Duppati *et al.*, 2020), whereas the other have found a negative relationship (Adams and Ferreira, 2009; Shehata *et al.*, 2017). Certain researchers have also suggested insignificant linkages between gender diversity and firm performance (Carter *et al.*, 2010; Martinez-Jimenez *et al.*, 2020).

As per Bennouri *et al.* (2018), women bring with them fresh viewpoints and diverse professional backgrounds on the board that differs from the "old boys' club". This enables them to give better advice to the managers (Anderson *et al.*, 2011). Women also act as proponents of positive behavior among the board members, which further enhances the monitoring role (Adams and Ferreira, 2009). The agency theory is one of the prime theories underlining the reduction in information asymmetry due to the presence of women directors (Reguera-Alvarado *et al.*, 2017), thereby leading to a substantial improvement in firm performance. Contrary to this, some studies have advocated that gender diversity can negatively impact financial performance owing to conflict in opinions and delay in decision-making in dynamic market scenarios where quick decisions are required (Santacreu-Vasut *et al.*, 2017). Individuals often perceive others based on social categories, such as gender and therefore, the gender-based categorization at the board level can impair the communication process as well as cooperation (Knippenberg *et al.*, 2004). Lastly, Green and Homroy (2018) suggest that for gender diversity to be effective, women should be appointed to workplaces where they can have decisional roles and can affect governance, without which the legal regulations and mandatory quotas would turn out to be fruitless. Moreover, the percentage of female representation on these boards has not reached the threshold where they can have a substantial say in the governance mechanisms and firm performance (Singh *et al.*, 2019; Martinez-Jimenez *et al.*, 2020). In the same vein, Garanina and Muravyev (2021) documented that any relation between women's representation and firm performance cannot be found until the representation reaches a threshold of three directors and similar results were corroborated when the number rose from a single women director to closer to the threshold limit (Nguyen *et al.*, 2015).

In the wake of mixed arguments, we formulate the following hypothesis:

H1. Gender diversity on boards has no impact on firm financial performance.

2.3 Gender diversity, board committees and financial performance

Board committees acting through directors are the major drivers for decision-making regarding the governance of the firms (Bugeja *et al.*, 2016) since they are vested with specific responsibilities of monitoring and advising the board (Chen and Wu, 2016). These committees accord several benefits such as knowledge specialization, decentralized decision-making, efficient work allocation, enhanced accountability of the board and mitigation of agency issues by allowing separation of management and control (Berezinets *et al.*, 2017). Prior research findings suggest that women directors are more inclined toward board monitoring (Gul *et al.*, 2011) and their presence enhances board effectiveness (Adams and Ferreira, 2009). This is based on the belief that an increase in women's representation sends a message of change in the governance mechanism (Wang and Lee, 2012) in the context of independent directors. But the literature is translucent in defining how this change in governance takes place. In the broader context, most of the studies have focused on finding a direct nexus between gender diversity on board and financial performance (Shehata *et al.*, 2017; Singh *et al.*, 2019; Duppati *et al.*, 2020; Brahma *et al.*, 2021). The results are typically representative of "tokenism" rather than representing the causal impact of gender diversity on performance (Green and Homroy, 2018).

However, the engagement and participation of women directors in these committees shall integrate the element of governance and bring out the benefits of female representation on boards (Green and Homroy, 2018). If women directors sit on these committees, the various advantages of diversity such as critical analysis of complex issues (Carter *et al.*, 2010) and risk averseness, are likely to improve firm performance (Kahane *et al.*, 2013). Based on the diversity of the committee members, the richness of the discussions shall also improve. Moreover, the board shall not be willing to appoint those directors who are not capable of making significant contributions to the board committees (Carter *et al.*, 2010). Ararat and Yurtoglu (2021) also found that when there is active involvement of women directors on board committees, meaningful results can be seen on firm value. As per them, such results are applicable only when women are seen on prominent committees and their representation reaches a certain threshold limit. Similar results were corroborated by Carter *et al.* (2010) and Green and Homroy (2018) in different institutional and contextual settings. Thus, we propose the following hypothesis:

H2. Gender diversity on the board committees has a positive impact on firm financial performance.

3. Data and methodology

3.1 Sample and data sources

To achieve the objectives of the study, we selected a sample of all Indian firms listed on the Standard and Poor (S&P) Bombay Stock Exchange (BSE) 500 index for a period of eight years, that is, from 2014 to 2021. The year 2014 has been taken as the starting point because of the amendment in the Indian Companies Act, in the year 2013, that mandated the appointment of at least one woman on board for certain classes of companies. The S&P BSE 500 index is a suitable measure because it represents almost 93% of the total market capitalization listed on the BSE. It comprises a diverse range of industries which shall accord richness to the industry-based outcomes (Raithatha and Haldar, 2021). Out of the sample, 35 banking and financial firms were removed and 37 firms with missing data were screened out, leaving a total of 428 firms and 3,424 firm-year observations.

The list of BSE 500 index and the relevant financial variables were sourced from the Prowess database, which is maintained by the Centre for Monitoring Indian Economy (CMIE). The Prowess database is a widely accepted database for retrieving data and conducting empirical studies based on the Indian corporate sector (Haldar *et al.*, 2018; Singh

et al., 2021a). The corporate governance variables were manually collected from the corporate governance reports of the sampled firms. We corrected and cross-verified many data points to avoid any inconsistencies that might be present in the data, using sources such as Bloomberg profile of directors, the Ministry of Corporate Affairs website and official and personal webpages of firms and directors to reconfirm the initials of the directors. This practice was followed based on the work of Haldar *et al.* (2020).

3.2 Measures of firm performance

We used two different variables to measure firm performance, both of which are used extensively and reliably in the literature. First is Tobin's *Q*, which is a prominent market-based measure of financial performance (Raithatha and Haldar, 2021; Sarkar and Selarka, 2021; Shehata *et al.*, 2017), and the second is the return on assets (ROA), an accounting-based measure that determines profitability (Duppati *et al.*, 2020; Ghalke *et al.*, 2022; Sarkar and Selarka, 2021).

Tobin's *Q* is defined as the market value of equity and debt divided by the total replacement cost of the assets. However, in many emerging economies, including India, most of the debt is not actively traded in the market and is held in the form of institutional debt. Moreover, the assets are also recorded at their historical cost in the books rather than the market value. Therefore, based on the previous literature (Sarkar and Selarka, 2021), Tobin's *Q* is calculated as:

$$Tobin's\ Q = (\text{Market Value of Equity} + \text{Book Value of Debt})/\text{Book Value of Assets}$$

Return on assets is calculated as:

$$ROA = \text{Net Profit before Interest, Tax, and Depreciation}/\text{Book Value of the Assets}$$

3.3 Measures of gender diversity

The major variable of our interest is gender diversity at the board level, on the remuneration committee and on the nomination committee. We employ three proxy variables to measure gender diversity. Firstly, we use the percentage of women directors (P-women; Ararat and Yurtoglu, 2021; Issa and Fang, 2019) measured as the ratio of the number of women directors on board/committee to the total number of directors on the board/committees. For robustness of results, we use two widely used diversity indices, Blau index and Shannon index (Issa and Fang, 2019; Shehata *et al.*, 2017; Singh *et al.*, 2019).

The Blau index and Shannon index are calculated as follows:

$$\text{Blau Index}: D = 1 - \sum_{i=1}^{N} p_i^2 \qquad \text{(Formula 1)}$$

$$\text{Shannon Index}: H' = -\sum_{i=1}^{S} p_I \ln p_i \qquad \text{(Formula 2)}$$

where.

P_i is the percentage of female directors in firm i

N and S are the numbers of different categories of gender in the firm.

Both indices consider the dissimilarity of the categories (i.e. male and female). The maximum and minimum value of the Blau index and Shannon index varies from 0 to 0.5 and 0 to 0.69, respectively (lowest to highest diversity). Both the indices are similar in their characteristics, except that Shannon index is more sensitive to change, being a logarithmic measure.

3.4 Control variables

Based on the past literature, we control for certain governance variables and firm-specific variables. In terms of governance variables, we control for board size, the number of independent directors, CEO duality and the number of board meetings held during the year. Based on the nexus between governance and firm performance, these variables are found to be impacting financial performance. The firm-specific parameters that affect the performance include firm size, leverage and organizational slack. Table 1 provides the details of the variables employed for the empirical analysis, along with references to past literature.

Variable	Definition	Supporting literature
(A) Dependent variables		
Tobin's *Q*	Market valuation of shares plus book value of debt divided by the book value of total assets	Nguyen *et al.* (2015), Shehata *et al.* (2017), Raithatha and Haldar (2021), Sarkar and Selarka (2021), Gupta *et al.* (2021a)
Return on asset (ROA)	Net Profit before interest, tax and depreciation divided by the total book value of assets employed by a firm	Duppati *et al.* (2020), Sarkar and Selarka (2021), Ghalke *et al.* (2022)
(B) Independent variables		
P-Women (percentage of women directors)	Percentage of women directors on board/committee in relation to the total number of directors on board/committee	Issa and Fang (2019), Ararat and Yurtoglu (2021), Singh *et al.* (2021a), Shukla and Teraiya (2022)
Blau index (Blau) (used for robustness)	$D = 1 - \sum_{i=1}^{N} p_i^2$	Shehata *et al.* (2017), Issa and Fang (2019), Singh *et al.* (2019)
Shannon index (Shannon) (used for robustness)	$H' = -\sum_{i=1}^{S} p_i \ln p_i$	Shehata *et al.* (2017), Issa and Fang (2019), Singh *et al.* (2019)
(C) Control variables		
(C.1) Governance variables		
Board size (B-size)	Number of directors present in the firm's board	Issa and Fang (2019), Pucheta-Martínez *et al.* (2019), Brahma *et al.* (2021), Mehrotra *et al.* (2021)
Number of independent directors (ID)	Number of independent directors on the board	Terjesen *et al.* (2016), Arora (2021)
CEO duality (duality)	If CEO and chairman of the board are the same, then 1, otherwise 0	Terjesen *et al.* (2016), Nguyen (2018), Bennouri *et al.* (2018), Hsu *et al.* (2021), Oware and Appiah (2022)
Number of board meetings (BM)	Number of board meetings held in the firm during a financial year	Terjesen *et al.* (2016), Ting *et al.* (2018), Al Farooque *et al.* (2020)
(C.2) Firm-specific variables		
Firm size (*F*-size)	Measured in terms of the natural log of total assets	Bennouri *et al.* (2018), Karim *et al.* (2019), Pucheta-Martínez *et al.* (2019)
Organizational slack (OS)	Calculated as the natural log of difference between current assets and current liabilities	Grüner and Raastad (2018), Lee *et al.* (2021)
Leverage (lever)	Calculated as the ratio of total long-term debt to the total assets	Ararat and Yurtoglu (2021), Karim *et al.* (2019), Pucheta-Martínez and Gallego-Álvarez (2020)

Source(s): The authors

Table 1. Summary of variables employed

3.5 Model specification and estimation

To find the association between the board as well as committee gender diversity and financial performance, the study uses the panel data regression method. The following econometric model is specified:

$$Financial_performance_{it} = \alpha + \beta_1 * gender\ diversity + \beta_2 * control_variables_{it} + \epsilon_{it} \quad (1)$$

where.

Financial Performance represents Tobin's *Q* and ROA.

Gender diversity represents Percentage of women directors, Blau and Shannon index.

Control variables represent board size, number of independent directors, CEO duality, number of board meetings, firm size, leverage and organizational slack.

ϵ_{it} represents the error term.

Two sets of regression equations are estimated based on two measures of financial performance, one using Tobin's *Q* and the other employing ROA. The vector of control variables represents the various governance and firm-related variables.

The panel data is a culmination of both cross-sectional as well as time-varying data, and various methods like pooled ordinary least squares, random effect and fixed-effect models are used to conduct the study. The panel data may suffer from various unobservable heterogeneity that needs to be addressed to avoid biased results. The endogeneity issue can be addressed either using the estimation technique such as the fixed or the random effects model or by using the instrumental variable in the analysis (Hermalin and Weisbach, 2003). To choose between the random effects model or fixed-effects model, the Hausman test was conducted (Hausman and Taylor, 1981). We found that the explanatory variable is correlated with the unobservable heterogeneity, and hence the regression results were conducted using the fixed-effects model. The methodology has been used based on some recent studies conducted in the context of gender diversity literature (Ararat and Yurtoglu, 2021; Akram *et al.*, 2020; Singh *et al.*, 2022).

4. Results and analysis

4.1 Descriptive statistics

Table 2 briefly presents the descriptive statistics, that is, mean, standard deviation, maximum values and minimum values, for our dependent, independent and control variables.

Variable	Mean	Standard deviation	Minimum	Maximum
Tobin's *Q*	3.90	3.27	0.213	73.78
ROA (%)	12.36	9.33	14.26	112.54
P-Women for Board	0.175	0.097	0.10	0.31
B-Size	11.53	3.54	3	54
ID	5.12	1.90	2	14
Duality	0.47	0.51	0	1
F-Size*(Rs. Millions)	4.89	0.977	0.93	8.13
Org-Slack*(Rs. Millions)	3.75	0.872	0.041	6.69
Lever	1.32	19.42	0	87
BM	6.62	2.96	6	63

Table 2. Descriptive statistics

Note(s): * denotes natural log
Source(s): The authors

The Tobin's *Q*, the market measure of financial performance, is found to have a mean value of 3.90, suggesting sound financial health of the sampled firms. The mean value of ROA is found to be 12.36 which means that on average, a firm can generate a return of 12.36% on the investments made, whereas the minimum and the maximum value are 14.26 and 112.54, respectively. The mean percentage of women on board is 0.104, indicating that the average representation of women on boards in our sample is merely 17.5%. The mean value of board size is found to be 11.53, and the minimum and maximum values are 3 and 54, respectively, representing that the majority of the companies have a large-sized board. The mean number of independent directors is 5.12 clearly stating that approximately 50% of the board is comprised of independent directors. The value of CEO duality is 0.47, indicating that in at least 50% of the boards, the position of CEO and the chairman is held by the same individual, thereby leading to a possibility of a high degree of control over decisions by the chairman. The firm size and organizational slack are found to have a mean value of 4.89 and 3.75 (in millions), representing the sound health of the BSE 500 firms. The mean value of leverage is 1.32 whereas the maximum and minimum values are found to be 87 and 0, respectively. The average number of board meetings conducted in a year is 6.62, which is reasonable.

4.2 Diagnostic tests

Prior to the hypothesis testing, it is pertinent to check the assumptions for panel data regression. Multicollinearity is tested using the variance inflation factor (VIF), tolerance values and Pearson correlation matrix. Multicollinearity may turn out to be a problem in case the value of VIF is greater than 10 and tolerance levels are near 0 (Gujarati and Porter, 2017) or if the values in the Pearson correlation matrix are more than 0.5. The correlation coefficients of all the variables, in our model for gender diversity on board, are less than 0.5 and the VIFs in the model vary from 1.02 to 2.82. The tolerance values also range from 0.46 to 0.99. Thus, we may conclude that there is no issue of multicollinearity in the model. Similar results are found in the model for gender diversity on the remuneration committee as well as the nomination committee.

Furthermore, it is crucial to monitor the heteroskedasticity and serial correlation. Indeed, if these are not corrected, the estimated variances, covariances and standard errors would be biased and inconsistent (Gujarati and Porter, 2017). Consequently, the Breusch–Pagan test is used to check for heteroskedasticity, and the Breusch–Godfrey test is used to detect autocorrelation. Both tests lead to the rejection of the null hypothesis, pointing toward their absence in our models (Wooldridge, 2015).

4.3 Impact of gender diversity on corporate boards and committees on financial performance (Tobin's Q)

Table 3 highlights the regression results obtained through the fixed effects model using Tobin's *Q* as a measure of financial performance. Three models have been created simultaneously, where Model 1 measures the percentage of women directors on board, Model 2 measures the percentage of women directors on remuneration committees, and Model 3 measures the percentage of women directors on the nomination committees. The coefficient values of P-women are found to be positive and significant at 1% for remuneration committee and nomination committee (Model 2, P-women = 0.383, *p*-value = 0.007; Model 3, P-women = 0.374, *p*-value = 0.010), but insignificant for the board (Model 1, P-women = 2.981, *p*-value = 0.241). The regression results support both H1 as well as H2 of our study and suggest that the presence of women directors on the prominent committees created by the board impacts financial performance through the participation of women in governance as well as the decision-making process (Green and Homroy, 2018).

Table 3.
Summary results of regression models using Tobin's *Q* as dependent variable

Variables	Model 1 (*p*-value)	Model 2 (*p*-value)	Model 3 (*p*-value)
P-Women for Board	2.981 (0.241)		
P-Women for Remuneration Committee		0.383*** (0.007)	
P-Women for Nomination Committee			0.374*** (0.010)
B-Size	0.0864** (0.017)	0.0671* (0.060)	0.0667* (0.062)
ID	−0.100* (0.092)	−0.0843 (0.162)	−0.0848 (0.160)
Duality	−0.373** (0.016)	−0.351** (0.026)	−0.350** (0.027)
BM	−0.0263 (0.368)	−0.0266 (0.371)	−0.0272 (0.359)
F-Size	−1.686*** (0.000)	−1.721*** (0.000)	−1.720*** (0.000)
OS	0.223 (0.144)	0.194 (0.212)	0.197 (0.206)
Leverage	−0.172*** (0.002)	−0.164*** (0.003)	−0.165*** (0.003)
Constant	9.896*** (0.000)	10.45*** (0.000)	10.45*** (0.000)

Note(s): *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$
Source(s): The authors

The results produce an interesting finding that integrating the female directors in the governance mechanisms by enhancing their participation in the prominent committees impacts financial performance, while merely putting them on the corporate board for the sake of on-paper diversity does not lead to any gains. Therefore, there lies a business case for gender diversity and encouragement of women's participation on board committees. The results are in line with the studies conducted in the context of emerging economies such as Turkey (Ararat and Yurtoglu, 2021) and European countries (Green and Homroy, 2018).

In terms of the control variables, board size is found to be positive and significant, implying that a larger board improves the financial performance of the firms. This justifies the resource dependency theory and is consistent with the previous studies (Adams and Mehran, 2012). The number of independent directors is found to be negatively correlated with the firm performance, though insignificant, based on the premise that independent directors may not have been able to bring the independence or diversity element in real terms. Instead, they may have added to additional costs in terms of hefty sitting fees (Arora and Sharma, 2016). The CEO duality is negatively associated with performance, as when the CEO and board chairman are the same person, there is a possibility of conflict of interest in decision-making, further hampering performance (Upadhyay and Zeng, 2014). An important control variable, that was expected to be positively associated with the financial performance but is found to be negative and insignificant, is the number of board meetings. This outcome may be attributed to the fact that very frequent meetings may negate their fruitfulness and instead add to the increased cost and time of the firm (Ting *et al.*, 2018; Chaudhary and Gakhar, 2018). The firm size is also found to be negatively associated with the firm performance, possibly because as firms grow, it becomes complicated to manage the various departments efficiently (Hamdan, 2018; Brahma *et al.*, 2021). Organizational slack is found to have a positive but insignificant impact on financial performance. Past studies have seen organizational slack as a positive dimension of performance and this is supported by organizational theory as well (Stock *et al.*, 2018). Lastly, the coefficient of leverage is negative and significant, indicating an inverse relationship between leverage and performance. This relation is based on two elements, the financial distress and the benefits based on the way debt has been serviced (González, 2013). In the Indian context, the costs have certainly outweighed the benefits.

4.4 Impact of gender diversity on corporate boards and committees on financial performance (return on assets)

Table 4 highlights the regression results using ROA as a measure of financial performance. Three simultaneous regression models were created to establish the relationship of various

Variables	Model 1 (*p*-value)	Model 2 (*p*-value)	Model 3 (*p*-value)
P-Women for Board	0.0188 (0.993)		
P-Women for Remuneration Committee		−0.217 (0.482)	
P-Women for Nomination Committee			−0.308 (0.323)
B-Size	0.0451* (0.060)	0.0470* (0.072)	0.0454* (0.076)
ID	0.267 (0.118)	0.258 (0.224)	0.256 (0.225)
Duality	−0.653* (0.052)	−0.595* (0.080)	−0.599* (0.078)
BM	−0.0452 (0.445)	−0.0656 (0.272)	−0.0666 (0.265)
F-Size	−5.875*** (0.000)	−5.912*** (0.000)	−5.904*** (0.000)
OS	3.049*** (0.000)	3.149*** (0.000)	3.153*** (0.000)
Leverage	−0.743 (0.290)	−0.741 (0.127)	−0.742 (0.343)
Constant	24.66*** (0.000)	24.71*** (0.000)	24.73*** (0.000)

Note(s): *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$
Source(s): The authors

Table 4. Summary results of regression models using ROA as dependent variable

variables of gender diversity on board and prominent committees with firm financial performance. The regression results in terms of the coefficient were found to be positive but insignificant in the case of gender diversity on board (Model 1, *P*-women = 0.0188, *p*-value = 0.993) but were negative and insignificant with respect to the prominent committees (Model 2, P-women = −0.217, *p*-value = 0.482; Model 3, P-women = −0.308, *p*-value = 0.323). Thus, both H1 and H2 are not supported.

The results concerning ROA contrast with the results in terms of Tobin's *Q*. The discrepancy between the two measures of financial performance underscores the fact that the accounting and market-based measures do not work on similar lines, and thus, there is a need to evaluate both the variables for the analysis (Bennouri *et al.*, 2018). The difference in the results arises because the appointment of female directors on board committees improves the perception of the investors (Green and Homroy, 2018) and other stakeholders which is readily reflected in the market-based performance measure, but it may take a longer time to get reflected in the accounting-based performance measure. The control variables produce similar results as in the case of Tobin's *Q*.

4.5 Robustness check

We further conduct a range of tests, which are aimed at enhancing the robustness of our results. Based on the extant corporate governance literature, this study uses P-women as the main independent variable to measure gender diversity. To check for the robustness of the findings of the study, two diversity indices have been used for gender diversity:

1. Blau index
2. Shannon index

When used as a proxy variable for gender diversity, the results are on the same lines as that of the main variables. Thus, the outcomes restate our point that the presence of women directors on prominent committees, and not just the board, improves the perception of the investors in terms of the firm's governance and thus gets reflected in the results, though the same is not the case with the accounting-based performance measure.

5. Discussions

There are two major insights that unfold from the findings of the study. First, merely putting the women on board to comply with the mandatory quotas may not lead to any visible benefits.

Since it is well established that the board performs most of its functions through prominent committees (such as remuneration, nomination and audit committees), making women a part of these committees and providing them with a platform to take up the responsibilities as well as voice their opinions has an indirect impact on the performance of the committees and the firms. This indirect impact accrues due to the active participation of women in strategic decisions and governance mechanisms which are assigned to these prominent committees.

Second, gender diversity on prominent committees significantly affects the market-based performance measure (Tobin's *Q*) but does not impact the accounting-based measure of performance (ROA). The plausible reason for the contradictory results could be attributed to the fact that firms with a representation of women on board committees are admired and are considered to be ethical firms, and therefore, they occupy a space in the good books of the society (Landry *et al.*, 2016). This perception is reflected in the market-based indicator, but the accounting results are based on the historical transactions and thus, may take some time to reflect the positive changes that accrue in terms of the governance. The results of the study further validate the theoretical framework in the form of agency theory and the resource dependency theory that provide support for gender diverse committees for the decision-making purpose.

India, as an emerging economy, is slowly picking up the pace of growth and development, not only on the economic front, but also on the social front. There are multiple constraints to the participation of women in economic, social and public domains. The institutional environment of India differs widely from various developed and developing economies based on its unique social and cultural setting. Due to such scenarios, the working environment for women is highly challenging in nature (Haq, 2012). In such a case, the benefits of their diverse experience, personality and perspective can only be taken when they are made part of prominent committees and are involved in the decision-making process. The above assumption got reflected in the outcomes of the study, where the gender diversity on board was found to be insignificant and it turned out significant in case of prominent committees.

Further there are two sets of theories, the similarity attraction theory (Bryne, 1971) as well as critical mass theory (Kanter, 1977a, b), that provide sound justification for insignificant results in case of gender diversity on board. According to similarity attraction theory (Bryne, 1971) individuals feel more comfortable interacting with others of the same gender (Chatman and O'Reilly, 2004) based on their likeliness and common characteristics. Similarly, critical mass theory (Kanter, 1977a, b) advocates the presence of at least three individuals to cause any influence on decision making. Thus, if more women directors are represented on the board, leading to critical mass, it could positively impacts board strategic tasks (Torchia *et al.*, 2011) which may result in better decision-making (Huse and Solberg, 2006) and performance.

6. Implications

The findings of the study offer various academic, policy, practical and managerial implications.

6.1 Academic and policy implications

The results of the study would help the future academicians explore further aspects of gender diversity and various committees on which these women are likely to be represented. It shall also open doors for initiating the discussions on the decision-making role played by women directors on various committees which, as of now, is an underexplored arena. Taking this study as a base, future researchers can delve into the board dynamics through a sentiment analysis, which can be useful to understand the subjective information in terms of what happens within the closed rooms of the board (Gupta *et al.*, 2020, 2021b).

The results of this study shall also help the policymakers and the regulators. This is a call for policymakers and regulatory bodies to act in a supervisory role and oversee the structure

and process for the appointment of women directors. This oversight may further give an upward push to the percentage of women directors and may lead to improved financial performance in the long run. Policymakers' and regulators' efforts should be focused on devising adequate provisions that could take advantage of diversity and provide an adequate critical mass of representation of women directors on the boards and committees.

6.2 Managerial and practical implications

As the gender diversity aspect is related to the management of the firms, the results of this study are very crucial for them. The corporate houses need to ensure that the presence of women on board as well as on committees is not merely an act of tokenism. Steps need to be taken to foster inclusivity to enhance the active participation of women in the decision-making process. Through such initiatives, the firms shall be able to leverage the benefits of gender diversity.

7. Conclusion

The aim of this study was to determine the relationship between gender diversity on corporate boards as well as on prominent board committees and the financial performance of firms, in an emerging economy. This relation was evaluated based on a sample of 3,424 firm-year observations from India. In the Indian context, we see the business case for promoting gender diversity based on the market-based results, however, there is still scope for wider participation of women on board and decision-making, which could be advanced once critical mass is achieved in the corporate boards. Our results suggest that gender diversity on board does not have any significant impact on financial performance, but gender diversity on prominent committees does have a bearing on the market-based performance indicator (Tobin's *Q*). Furthermore, the outcomes suggest that benefits of gender diversity can accrue to the firms only when these women sit on prominent committees and are engaged in the strategic responsibilities and duties assigned to these committees.

India, an emerging economy, has a unique set of cultural and hierarchical features driven by a deep-rooted patriarchal mindset. Despite the constraints on women's participation in the workforce, we find a positive and significant relationship between women's representation on prominent committees and firm financial performance.

7.1 Limitations and future scope

Our study has certain limitations which provide scope for future research. Firstly, the study is based on just one emerging economy that is India. Future studies could be conducted using multiple developing and emerging economies which shall be more insightful in terms of highlighting the cultural and demographic landscape. Secondly, we focused on the performance of only large, listed firms which form part of the S&P BSE 500 index. In the future, small and medium firms can be considered for better and wholesome generalization of the results along with the non-listed firms. Thirdly, the study did not focus on the degree of diversity by making a categorization between firms that have one female director and more than one female director. The above dimension is likely to provide interesting insights. The future scholars may also explore the various types of gender stereotypes that exist in the society and how can these stereotypes be minimized for a better future and wellbeing.

References

Abi Rached, M., Hankir, A. and Zaman, R. (2021), "Emotional abuse in women and girls mediated by patriarchal upbringing and its impact on sexism and mental health: a narrative review", *Psychiatria Danubina*, Vol. 33, pp. 137-144.

Adams, R.B. and Ferreira, D. (2009), "Women in the boardroom and their impact on governance and performance", *Journal of Financial Economics*, Vol. 94 No. 2, pp. 291-309, doi: 10.1016/j.jfineco.2008.10.007.

Adams, R.B. and Mehran, H. (2012), "Bank board structure and performance: evidence for large bank holding companies", *Journal of Financial Intermediation*, Vol. 21 No. 2, pp. 243-267.

Adams, R.B., Hermalin, B.E. and Weisbach, M.S. (2010), "The role of board of directors in corporate governance", *Journal of Economic Literature*, Vol. 48 No. 1, pp. 58-107.

Ain, Q.U., Yuan, X., Javaid, H.M., Usman, M. and Haris, M. (2021), "Female directors and agency costs: evidence from Chinese listed firms", *International Journal of Emerging Markets*, Vol. 16 No. 8, pp. 1604-1633, doi: 10.1108/IJOEM-10-2019-0818.

Akram, F., Abrar ul Haq, M., Natarajan, V.K. and Chellakan, R.S. (2020), "Board heterogeneity and corporate performance: an insight beyond agency issues", *Cogent Business and Management*, Vol. 7 No. 1, pp. 1-14.

Al Farooque, O., Buachoom, W. and Sun, L. (2020), "Board, audit committee, ownership and financial performance – emerging trends from Thailand", *Pacific Accounting Review*, Vol. 32 No. 1, pp. 54-81, doi: 10.1108/PAR-10-2018-0079.

Ames, D.A., Hines, C.S. and Sankara, J. (2018), "Board risk committees: insurer financial strength ratings and performance", *Journal of Accounting and Public Policy*, Vol. 37 No. 2, pp. 130-145.

Anderson, R.C., Reeb, D.M., Upadhyay, A. and Zhao, W. (2011), "The economics of director heterogeneity", *Financial Management*, Vol. 40 No. 1, pp. 5-38.

Ararat, M. and Yurtoglu, B.B. (2021), "Female directors, board committees, and firm performance: time-series evidence from Turkey", *Emerging Markets Review*, Vol. 48, 100768.

Arora, A. (2021), "Gender diversity in boardroom and its impact on firm performance", *Journal of Management and Governance*, Vol. 25 No. 1, pp. 1-21.

Arora, A. and Sharma, C. (2016), "Corporate governance and firm performance in developing countries: evidence from India", *Corporate Governance*, Vol. 16 No. 2, pp. 420-436, doi: 10.1108/CG-01-2016-0018.

Baker, H.K., Pandey, N., Kumar, S. and Haldar, A. (2020), "A bibliometric analysis of board diversity: current status, development, and future research directions", *Journal of Business Research*, Vol. 108, pp. 232-246.

Bennouri, M., Chtioui, T., Nagati, H. and Nekhili, M. (2018), "Female board directorship and firm performance: what really matters?", *Journal of Banking and Finance*, Vol. 88, pp. 267-291.

Berezinets, I., Ilina, Y. and Cherkasskaya, A. (2017), "Board structure, board committees and corporate performance in Russia", *Managerial Finance*, Vol. 43 No. 10, pp. 1073-1092, doi: 10.1108/MF-11-2015-0308.

Brahma, S., Nwafor, C. and Boateng, A. (2021), "Board gender diversity and firm performance: the UK evidence", *International Journal of Finance and Economics*, Vol. 26 No. 4, pp. 5704-5719.

Bryne, D. (1971), *The Attraction Paradigm*, Academic Press, New York.

Bugeja, M., Matolcsy, Z. and Spiropoulos, H. (2016), "The association between gender-diverse compensation committees and CEO compensation", *Journal of Business Ethics*, Vol. 39 No. 2, pp. 375-390.

Carter, D.A., Simkins, B.J. and Simpson, W.G. (2003), "Corporate governance, board diversity, and firm value", *Financial Review*, Vol. 38 No. 1, pp. 33-53.

Carter, D.A., D'Souza, F., Simkins, B.J. and Simpson, W.G. (2010), "The gender and ethnic diversity of US boards and board committees and firm financial performance", *Corporate Governance: An International Review*, Vol. 18 No. 5, pp. 396-414.

Chatman, J.A. and O'Reilly, C.A. (2004), "Asymmetric reactions to work group sex diversity among men and women", *Academy of Management Journal*, Vol. 47 No. 2, pp. 193-208, doi: 10.2307/20159572.

Chaudhary, N. and Gakhar, K. (2018), "Corporate governance and financial performance with a perspective on board size and frequency of board meetings: empirical evidence from India", *Drishtikon: A Management Journal*, Vol. 9 No. 1, p. 37.

Chen, M.Y. and Kao, C.L. (2020), "Women on boards of directors and firm performance: the mediation of employment downsizing", *The International Journal of Human Resource Management*, Vol. 33 No. 13, pp. 2597-2629.

Chen, K.D. and Wu, A. (2016), *The Structure of Board Committees*, Vol. 1, Harvard Business School, Boston, MA.

Credit Suisse (2021), "The CS Gender 3000 in 2021: broadening the diversity discussion", available at: https://www.bing.com/ck/a?!&&p=66f8e08a2ab3539a109b982673061ac4758e29c525ed349730356c604597abf1JmltdHM9MTY0Nzg2MDE3MCZpZ3VpZD1mMDZmNjI2MC0zNDMyLTQ0YzQtOGQzZC02OWFmZTYyOWY5OTEmaW5zaWQ9NTE0Ng&ptn=3&fclid=84563b27-a905-11ec-a8df-03926ebbe2eb&u=a1aHR0cHM6Ly93d3cuY3JlZGl0LXN1aXNzZS5jb20vbWVkaWEvYXNzZXRzL2NvcnBvcmF0ZS9kb2NzL2Fib3V0LXVzL3Jlc2VhcmNoL3B1YmxpY2F0aW9ucy9jc3JpLTIwMjEtZ2VuZGVyLTMwMDAucGRmP21zY2xraWQ9ODQ1NjNiMjdhOTA1MTFlY2E4ZGYwMzkyNmViYmUyZWI&ntb=1 (accessed 4 January 2022).

Duppati, G., Rao, N.V., Matlani, N., Scrimgeour, F. and Patnaik, D. (2020), "Gender diversity and firm performance: evidence from India and Singapore", *Applied Economics*, Vol. 52 No. 14, pp. 1553-1565.

Eckes, T. and Trautner, H.M. (2012), *Developmental Social Psychology of Gender: An Integrative Framework*, New York, Psychology Press, pp. 17-46.

Feng, M., Kang, T. and Nabar, S. (2017), "National societal values and corporate governance", *International Journal of Emerging Markets*, Vol. 12 No. 2, pp. 183-198, doi: 10.1108/IJoEM-08-2015-0150.

Garanina, T. and Muravyev, A. (2021), "The gender composition of corporate boards and firm performance: evidence from Russia", *Emerging Markets Review*, Vol. 48, 100772.

Ghalke, A., Haldar, A. and Kumar, S. (2022), "Family firm ownership and its impact on performance: evidence from an emerging market", *Review of Managerial Science*, pp. 1-20, doi: 10.1007/s11846-022-00527-7.

González, V.M. (2013), "Leverage and corporate performance: international evidence", *International Review of Economics and Finance*, Vol. 25, pp. 169-184.

Green, C.P. and Homroy, S. (2018), "Female directors, board committees and firm performance", *European Economic Review*, Vol. 102, pp. 19-38.

Groysberg, B. and Bell, D. (2013), "Dysfunction in the boardroom", *Harvard Business Review*, Vol. 91, pp. 89-97, available at: https://www.inspiringcapital.ly/wp-content/uploads/2018/08/Dysfunction-in-theBoardroom.pdf.

Grüner, A. and Raastad, I. (2018), "Financial slack and firm performance during economic downturn", *GSL Journal of Business Management and Administrative Affairs*, Vol. 1 No. 1, pp. 1-7.

Gujarati, D.N. and Porter, D.C. (2017), *Basic Econometrics*, McGrew Hill Book.

Gul, F.A., Srinidhi, B. and Ng, A.C. (2011), "Does board gender diversity improve the informativeness of stock prices?", *Journal of Accounting and Economics*, Vol. 51 No. 3, pp. 314-338.

Gupta, N. (2017), "Gender inequality in the work environment: a study of private research organizations in India", *Equality, Diversity and Inclusion*, Vol. 36 No. 3, pp. 255-276, doi: 10.1108/EDI-04-2016-0029.

Gupta, V., Singh, S. and Yadav, S. (2019), "Impact of anchor investors on IPO returns during pre-market and aftermarket: evidence from India", *Journal of Advances in Management Research*, Vol. 17 No. 3, pp. 351-368.

Gupta, V., Singh, S. and Yadav, S.S. (2020), "IPO under-pricing and media sentiments: avenue supermarts and Vaswani industries", *South Asian Journal of Business and Management Cases*, Vol. 9 No. 2, pp. 178-188.

Gupta, L., Sardana, V., Narula, G. and Singhania, S. (2021a), "Does corporate social responsibility affect corporate financial performance? A Myth or a reality", *Business Analyst*, Vol. 42 No. 1, pp. 71-100.

Gupta, V., Singh, S. and Yadav, S.S. (2021b), "The impact of media sentiments on IPO under-pricing", *Journal of Asia Business Studies*, Vol. ahead-of-print No. ahead-of-print, doi: 10.1108/JABS-10-2020-0404.

Haldar, A., Shah, R., Nageswara Rao, S.V.D., Stokes, P., Demirbas, D. and Dardour, A. (2018), "Corporate performance: does board independence matter? – Indian evidence", *International Journal of Organizational Analysis*, Vol. 26 No. 1, pp. 185-200, doi: 10.1108/IJOA-12-2017-1296.

Haldar, A., Datta, S. and Shah, S. (2020), "Tokenism or realism? Gender inclusion in corporate boards", *Equality, Diversity and Inclusion*, Vol. 39 No. 6, pp. 707-725, doi: 10.1108/EDI-04-2019-0126.

Hamdan, A. (2018), "Intellectual capital and firm performance: differentiating between accounting-based and market-based performance", *International Journal of Islamic and Middle Eastern Finance and Management*, Vol. 11 No. 1, pp. 139-151, doi: 10.1108/IMEFM-02-2017-0053.

Haq, R. (2012), "The managing diversity mindset in public versus private organizations in India", *The International Journal of Human Resource Management*, Vol. 23 No. 5, pp. 892-914, doi: 10.1080/09585192.2012.651297.

Hausman, J.A. and Taylor, W.E. (1981), "Panel data and unobservable individual effects", *Econometrica: Journal of the Econometric Society*, Vol. 49 No. 6, pp. 1377-1398.

Hermalin, B.E. and Weisbach, M.S. (2003), "Boards of directors as an endogenously determined institution: a survey of the economic literature", *FRB New York - Economic Policy Review 2003*, Vol. 9, 1, Apr, pp. 7-26.

House, R.J., Hanges, P.J., Javidan, M., Dorfman, P. and Gupta, V. (2004), *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*, Sage Publications, Thousand Oaks, CA.

Hsu, S., Lin, S.W., Chen, W.P. and Huang, J.W. (2021), "CEO duality, information costs, and firm performance", *The North American Journal of Economics and Finance*, Vol. 55, 101011.

Hubner, S., Frese, M., Song, Z., Tripathi, N., Kaschner, T. and Le Kong, X. (2022), "An Asia-centric approach to team innovation: cultural differences in exploration and exploitation behavior", *Journal of Business Research*, Vol. 138, pp. 408-421.

Huse, M. and Solberg, A.G. (2006), "Gender-related boardroom dynamics: how Scandinavian women make and can make contributions on corporate boards", *Women in Management Review*, Vol. 21 No. 2, pp. 113-130, doi: 10.1108/09649420610650693.

International Monetary Fund (2021), "World economic Outlook report", available at: https://www.imf.org/en/Publications/WEO (accessed 12 February 2022).

Issa, A. and Fang, H.X. (2019), "The impact of board gender diversity on corporate social responsibility in the Arab Gulf states", *Gender in Management: An International Journal*, Vol. 34 No. 7, pp. 577-605, doi: 10.1108/GM-07-2018-0087.

Jayachandran, S. (2021), "Social norms as a barrier to women's employment in developing countries", *IMF Economic Review*, Vol. 69 No. 3, pp. 576-595.

Jensen, M.C. and Meckling, W.H. (1976), "Theory of the firm: managerial behaviour, agency costs and ownership structure", *Journal of Financial Economics*, Vol. 3 No. 4, pp. 305-360.

Jha, S.K. and Jha, S. (2022), "Gender discrimination and retaliation at pinterest", *FIIB Business Review*, p. 231971452210996.

Kahane, L., Longley, N. and Simmons, R. (2013), "The effects of coworker heterogeneity on firm-level output: assessing the impacts of cultural and language diversity in the National Hockey League", *Review of Economics and Statistics*, Vol. 95 No. 1, pp. 302-314.

Kanter, R. (1977a), *Men and Women of the Organization*, Basic Books, New York, pp. 206-242.

Kanter, R. (1977b), "Some effects of proportions on group life: skewed sex ratios and responses to token women", *American Journal of Sociology*, Vol. 82 No. 5, pp. 965-990.

Karim, S., Manab, N.A. and Ismail, R.B. (2019), "The dynamic impact of board composition on CSR practices and their mutual effect on organizational returns", *Journal of Asia Business Studies*, Vol. 14 No. 4, pp. 463-479, doi: 10.1108/JABS-07-2019-0214.

Knippenberg, D.V., De Dreu, C.K. and Homan, A.C. (2004), "Work group diversity and group performance: an integrative model and research agenda", *Journal of Applied Psychology*, Vol. 89 No. 6, p. 1008.

Kumar, R. (2020), *Confessional Poetry*, Orange Books Publication.

Landry, E.E., Bernardi, R.A. and Bosco, S.M. (2016), "Recognition for sustained corporate social responsibility: female directors make a difference", *Corporate Social Responsibility and Environmental Management*, Vol. 23 No. 1, pp. 27-36, doi: 10.1002/csr.1358.

Lattemann, C. (2014), "On the convergence of corporate governance practices in emerging markets", *International Journal of Emerging Markets*, Vol. 9 No. 2, pp. 316-332, doi: 10.1108/IJoEM-06-2013-0093.

Lee, T., Liu, W.T. and Yu, J.X. (2021), "Does TMT composition matter to environmental policy and firm performance? The role of organizational slack", *Corporate Social Responsibility and Environmental Management*, Vol. 28 No. 1, pp. 196-213.

Maji, S.G. and Saha, R. (2021), "Gender diversity and financial performance in an emerging economy: empirical evidence from India", *Management Research Review*, Vol. 44 No. 12, pp. 1660-1683, doi: 10.1108/MRR-08-2020-0525.

Martinez-Jimenez, R., Hernández-Ortiz, M.J. and Cabrera Fernández, A.I. (2020), "Gender diversity influence on board effectiveness and business performance", *Corporate Governance*, Vol. 20 No. 2, pp. 307-323, doi: 10.1108/CG-07-2019-0206.

Martinsons, M.G. and Davison, R.M. (2007), "Strategic decision making and support systems: comparing American, Japanese and Chinese managemen", *Decision Support Systems*, Vol. 43 No. 1, pp. 284-300.

Mehrotra, S., Mohanty, B. and Sharma, T. (2021), "Do board quality and promoters' holdings affect firm performance? Evidence from small and medium-sized enterprises", *FIIB Business Review*, p. 231971452098028.

Nguyen, T.X.T. (2018), "Corporate governance and conglomerate diversification strategy – evidence from Vietnam", *International Journal of Emerging Markets*, Vol. 13 No. 6, pp. 1578-1596, doi: 10.1108/IJoEM-10-2016-0260.

Nguyen, T., Locke, S. and Reddy, K. (2015), "Does boardroom gender diversity matter? Evidence from a transitional economy", *International Review of Economics and Finance*, Vol. 37, pp. 184-202.

Nicolò, G., Sannino, G. and Iorio, S.D. (2021), "Gender diversity and online intellectual capital disclosure: evidence from Italian-listed firms", *Journal of Public Affairs*, e2706.

Oware, K.M. and Appiah, K. (2022), "CEO characteristics, family-controlled firms and employer-support volunteerism: evidence from India", *FIIB Business Review*, p. 231971452211059.

Pfeffer, J. and Salancik, G.R. (1978), *The External Control of Organisations: A Resource Dependence Perspective*, Harper & Row, New York, NY.

Pucheta-Martínez, M.C. and Gallego-Álvarez, I. (2020), "Corporate environmental disclosure practices in different national contexts: the influence of cultural dimensions", *Organization and Environment*, Vol. 33 No. 4, pp. 597-623, doi: 10.1177/1086026619860263.

Pucheta-Martínez, M.C., Bel-Oms, I. and Olcina-Sempere, G. (2019), "Commitment of independent and institutional women directors to corporate social responsibility reporting", *Business Ethics: A European Review*, Vol. 28 No. 3, pp. 290-304, doi: 10.1111/beer.12218.

Raithatha, M. and Haldar, A. (2021), "Are internal governance mechanisms efficient? The case of a developing economy", *IIMB Management Review*, Vol. 33 No. 3, pp. 191-204.

Rebérioux, A. and Roudaut, G. (2017), "Gender quota and inequalities inside the boardroom", École Polytechnique Université Paris-Saclay, available at: https://hal-polytechnique.archives-ouvertes.fr/hal-01618949.

Reguera-Alvarado, N., De Fuentes, P. and Laffarga, J. (2017), "Does board gender diversity influence financial performance? Evidence from Spain", *Journal of Business Ethics*, Vol. 141 No. 2, pp. 337-350, doi: 10.1007/s10551-015-2735-9.

Santacreu-Vasut, E., Shenkar, O. and Shoham, A. (2017), "Linguistic gender marking and its international business ramifications", *Language in International Business*, Palgrave Macmillan, Cham, Vol. 45, pp. 1170-1178.

Sarkar, J. and Selarka, E. (2021), "Women on board and performance of family firms: evidence from India", *Emerging Markets Review*, Vol. 46, 100770.

Schwartz, S.H. (2014), "Rethinking the concept and measurement of societal culture in light of empirical findings", *Journal of Cross-Cultural Psychology*, Vol. 45 No. 1, pp. 5-13.

Shahroor, H.G. and Ismail, A.I. (2022), "Corporate governance mechanisms and earnings management activities: evidence from the UAE banking sector", *FIIB Business Review*, p. 231971452210990.

Shehata, N., Salhin, A. and El-Helaly, M. (2017), "Board diversity and firm performance: evidence from the UK SMEs", *Applied Economics*, Vol. 49 No. 48, pp. 4817-4832, doi: 10.1080/00036846.2017.1293796.

Shukla, H. and Teraiya, V. (2022), "Influence of gender diversity in boards on family and non-family businesses towards innovation and creativity", *FIIB Business Review*, p. 231971452110686.

Singh, S.B. (2021), "Delhi high court expresses worry over alarming increase of false rape cases", *The Hindu*, 22 August, available at: https://www.thehindu.com/news/national/delhi-high-court-expresses-worry-over-alarming-increase-of-false-rape-cases/article36042093.ece.

Singh, A.K., Singhania, S. and Sardana, V. (2019), "Do women on boards affect firm's financial performance? Evidence from Indian IPO firms", *Australasian Accounting, Business and Finance Journal*, Vol. 13 No. 2, pp. 53-68, doi: 10.14453/aabfj.v13i2.4.

Singh, A.K., Kota, H.B., Sardana, V. and Singhania, S. (2021a), "Does gender diversity on board promote corporate social responsibility? An empirical analysis of sustainable development goals", *Australasian Accounting, Business and Finance Journal*, Vol. 15 No. 5, pp. 22-40.

Singh, J., Singhania, S. and Aggrawal, D. (2021b), "Gender diversity on corporate boards: review and future research agenda through bibliometric mapping", *Corporate Governance and Sustainability Review*, Vol. 5 No. 3, pp. 57-72, doi: 10.22495/cgsrv5i3p5.

Singh, J., Singhania, S. and Aggrawal, D. (2022), "Does board gender diversity impact financial performance? Evidence from the Indian IT sector", *Society and Business Review*, ahead-of-print.

Stock, G., Greis, N. and Fischer, W. (2018), "Organisational slack and new product time to market performance", *International Journal of Innovation Management*, Vol. 22 No. 4, 1850034.

Terjesen, S., Couto, E.B. and Francisco, P.M. (2016), "Does the presence of independent and female directors impact firm performance? A multi-country study of board diversity", *Journal of Management and Governance*, Vol. 20 No. 3, pp. 447-483.

Ting, I.W.K., Kweh, Q.L. and Hoanh, L.T.H. (2018), "Board meeting frequency and financial performance: a case of listed firms in Vietnam", *International Journal of Business and Society*, Vol. 19 No. 2, pp. 464-472.

Torchia, M., Calabrò, A. and Huse, M. (2011), "Women directors on corporate boards: from tokenism to critical mass", *Journal of Business Ethics*, Vol. 102 No. 2, pp. 299-317, doi: 10.1007/s10551-011-0815-z.

Tripathi, R. and Vijayan, U. (2020), "Power distance: an Indian perspective", available at: https://www.iimb.ac.in/turn_turn/power-distance-indian-perspective.php (accessed 10 December 2021).

Upadhyay, A. and Zeng, H. (2014), "Gender and ethnic diversity on boards and corporate information environment", *Journal of Business Research*, Vol. 67 No. 11, pp. 2456-2463.

Wang, M.C. and Lee, Y.C. (2012), "The signalling effect of independent director appointments", *Emerging Markets Finance and Trade*, Vol. 48 No. 5, pp. 25-47.

Wooldridge, J.M. (2015), "Control function methods in applied econometrics", *Journal of Human Resources*, Vol. 50 No. 2, pp. 420-445.

World Economic Forum (2021), "Global gender gap report 2021", available at: https://www.weforum.org/reports/global-gender-gap-report-2021 (accessed 14 February 2022).

Corresponding author
Jagvinder Singh can be contacted at: singhjagvinder727@gmail.com

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

Journal of Physics D: Applied Physics

ACCEPTED MANUSCRIPT

Carrier Concentration Mediated Enhancement in Thermoelectric Performance of Various Polymorphs of Hafnium Oxide: A Plausible Material for High Temperature Thermoelectric Energy Harvesting Application

To cite this article before publication: Rajesh Kumar *et al* 2022 *J. Phys. D: Appl. Phys.* in press https://doi.org/10.1088/1361-6463/ac9986

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is © **2022 IOP Publishing Ltd**.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.
As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by-nc-nd/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 14.139.251.98 on 25/10/2022 at 06:09

Carrier Concentration Mediated Enhancement in Thermoelectric Performance of Various Polymorphs of Hafnium Oxide: A Plausible Material for High Temperature Thermoelectric Energy Harvesting Application

Rajesh Kumar [a], Ramesh Kumar [b], Mukhtiyar Singh [a †], Deshraj Meena [a] and Ankush Vij [c]

[a] Department of Applied Physics, Delhi Technological University, Delhi-110042, India

[b] Department of Physics, Guru Jambheshwar University of Science & Technology, Hisar-125001, India

[c] Department of Physics, University of Petroleum and Energy Studies, Uttarakhand-248001, India

[†] msphysik09@gmail.com, mukhtiyarsingh@dtu.ac.in

Abstract: The optimization of figure of merit by tuning carrier concentrations is an effective way to realize efficient thermoelectrics. Recently, the feasibility of high *p*-type carrier concentration (order of ~10^{22} cm^{-3}) is experimentally demonstrated in various polymorphs of HfO_2. In light of these studies, using the first-principles calculation combined with the semi-classical Boltzmann transport theory and phonon dynamics, we realized high thermoelectric performance in various polymorphs of HfO_2 in a range of carrier concentrations at high temperatures. The phonon dispersion calculations confirm the dynamical stability of all polymorphs. The observed values of the Seebeck coefficient are 945.27 mV/K, 922.62 mV/K, 867.44 mV/K, and 830.81 mV/K for tetragonal (*t)*, orthorhombic (*o*), monoclinic (*m)*, and cubic (*c)* phases of HfO_2, respectively, at 300 K. These values remain positive at all studied temperatures which ensures the *p*-type behaviour of HfO_2 polymorphs. The highest value of electrical conductivity (2.34×10^{20} $\Omega^{-1}m^{-1}s^{-1}$) observed in *c*-HfO_2 at 1200 K, and the lowest value of electronic thermal conductivity (0.37×10^{15} W/mKs) observed in *o*-HfO_2 at 300 K. The lattice thermal conductivities at room temperature are 5.56 W/mK, 2.87 W/mK, 4.32 W/mK, and 1.75 W/mK for *c*-, *m*-, *o*- and *t*- HfO_2, respectively which decrease to 1.58 W/mK, 0.92 W/mK, 1.12 W/mK, 0.53 W/mK at 1200 K for respective phases. The low lattice thermal conductivities lead to the high values of the figure of merit, i.e., 0.97, 0.87, 0.83, and 0.77 at 1200 K for the *m*-, *o*-, *t*-, and *c*- HfO_2, respectively, at the optimized carrier concentrations (~10^{21} cm^{-3}). The predicted optimized carrier concentrations for various phases are in close agreement with the experimental reports. The estimated high figure of merit can make HfO_2 a potential material for thermoelectric energy harvesting applications at elevated temperatures.

Accepted Manuscript

Keywords: Density Functional Theory; BoltzTrap; HfO_2; Thermoelectrics; Lattice thermal conductivity

1. Introduction:

The unprecedented increasing demand of energy worldwide has wrecked the environment irreparably leading to the devastating problem of global warming. This has led to new avenues of research toward green and clean energy alternatives [1-2]. The conventional energy resources such as wind, geothermal, and solar have their own constraints regarding location and weather [3]. Thermoelectric (TE) energy conversion, on the other hand, has emerged as a potential technique to convert waste heat from automobiles, factories, etc., to useful electricity [4]. This technique is foreseen as a clean and sustainable future energy source, given that these TE devices are eco-friendly, silent, and reliable [5]. The performance of a TE material is usually gauged by a dimensionless quantity known as the figure of merit and given as $ZT = \sigma S^2 T/K_{total}$, where K_{total} (contribution of electronic thermal conductivity (K_e) and lattice thermal conductivity (K_l)) represent the total thermal conductivity, T is temperature, σ is electrical conductivity, and S is Seebeck coefficient [6]. To realize high value of ZT, a TE material indeed has a low value of K_{total} and a high value σ and S [7]. Augmentation in TE performance of a material is significantly contributed by phonon and carrier transport [6,8]. Some traditional TE materials like Bi_2Te_3, Si_2Te_3, and PbTe have the drawback of low ZT at a high temperature range [9-10]. To achieve efficient TE performance at high temperatures, a variety of oxide-based materials, including some wide band gap semiconductors, have been theoretically predicted and experimentally materialized [11-15].

Recently, it has been reported that various kinds of doping in Hafnium oxide (HfO_2) could enhance the ZT [16-19]. The high thermal stability [20], high melting point, and mechanical resistance [21] of HfO_2 make it a prominent host for TE applications. The synthesis temperatures vary significantly for different polymorphs of HfO_2. A low temperature thermodynamically stable monoclinic (*m*) phase exists at ~1200 ^oC, whilst tetragonal (*t*) phase and cubic (*c*) phase occurs at 1700 ^oC and 2500 ^oC [22-24]. The high pressure transforms the *m*-HfO_2 to orthorhombic (*o*) phase [25]. The existence of a variety of polymorphic phases of HfO_2 has numerous industrial applications-like ferroelectric material [26], scintillation counter [27], and resistive random access memory devices [28].

Although, the pure HfO_2 has a wide band gap of ~5.7 eV which is too large in the sense of thermoelectric applications but it naturally contains large concentrations of oxygen defects

Accepted Manuscript

which are generated during the sample growth and deposition. It has been observed that these oxygen vacancies play a key role in upsurging the carrier concentration of various polymorphs of HfO_2 and in reducing the band gap which make it an excellent thermoelectric at high temperature. For instance, Hilderbrandt et al. have reported that *p*-type carrier concentration of the order 6×10^{21} cm^{-3} in HfO_{2-x} and reduction of band gap from 5.7 eV to 4.5 eV have been observed [29]. In a similar study, Kaiser et al. have found carrier concentration varies from 3.4×10^{21} cm^{-3} to 1.8×10^{22} cm^{-3}, while Hadacek et al. found 10^{22} cm^{-3} *p*-type charge carrier concentrations, in HfO_2 [30,31]. Kumar et al. have demonstrated the reduction in band gap of *m*-HfO_2 owing to oxygen defects [21]. In addition to these studies, band engineering has been established as an effective technique to reduce the band gap and enhancing carrier concentration. González et al. have reported that the band engineering via Mn-doing in *m*-HfO_2 results in band gap reduction up to 2.1 eV [32]. Perevalov et al have realized the carrier concentration of 8.7×10^{21} cm^{-3} with the co-doping of Zr and La in HfO_2 [33]. In a theoretical investigation, Islamov et al have demonstrated the carrier concentration of the order 10^{21} cm^{-3} in Zr doped *o*-HfO_2 with oxygen defect [34].

Besides enhancing carrier concentration and reducing band gap, the oxygen vacancies may also help in reducing the thermal conductivity in metal oxides [35-37]. Various reports of relatively low K_l for HfO_2 over other metal oxides have existed in the literature [38-39]. The possibility of high carrier concentrations with naturally occurring oxygen defects and relatively low lattice thermal conductivity are two major attributes that can make HfO_2 an excellent thermoelectric energy harvesting material. Hence, it is vital to evaluate the effect of carrier concentration and the contribution of lattice dynamics in thermal conductivity for true estimation of *ZT*.

Here, we explored the various polymorphs of HfO_2 in a broad temperature range (300 K -1200 K) at varying carrier concentrations to achieve optimum *ZT*. The detailed investigation of structural, electronic, phonon transport, and thermoelectric properties is carried out using first-principles method and semi-classical Boltzmann transport theory. The phonon behaviour is examined to check the dynamical stability of the studied systems and to evaluate the lattice thermal conductivities. We demonstrated how the optimized carrier concentration can enhance the thermoelectric performance of the materials under investigation. To the best of our knowledge no such study, including lattice thermal conductivities of various polymorphs of HfO_2, has been yet reported.

2. Computational Methodology:

Accepted Manuscript

The electronic and transport properties were evaluated using the formalism of density functional theory (DFT) as implemented in WIEN2K code (a full potential linearized augmented plane wave method) [40]. The generalized gradient approximation (GGA) scheme for exchange and correlation (XC) potentials, given by Perdew-Burke-Ernzerhof (PBE), was used for structural optimization [41]. The approximations such as local density approximation and GGA underestimate the band gap [42]. Therefore, we have used Tran-Blaha modified Becke-Johnson (TB-mBJ) approximation [43] to comprehend exchange and correlation potentials which uses semilocal quantities to reproduce well step structures. This has been tested on several wide and small band gap materials like SnO_2, LiCl, GeO_2, NiO, FeO, MgO etc. and the obtained value of band gaps are in agreement with the experimental results [43]. The other functionals such HSE06 and GW approximate give similar value of band gap, as predicted by TB-mBJ for HfO_2, but are more costly and time consuming [44]. Therefore, we have chosen TB-mBJ for the presented study.

For precise calculations, a K-mesh of 20×20×20 with 256 K-points were used in the Irreducible Brillouin Zone (IBZ) for all the polymorphs of HfO_2. The plane wave cutoff was decided by the $R_{MT}Kmax = 7$, and the muffin tin sphere radii were chosen to be non-overlapping with RMT of 2.18 for Hf atom and 1.5 for O atom. The criterion for energy convergence was set to 0.0001 Ry/cell and the cut-off energy of core and valence state was fixed at -6 Ry. The temperature dependent thermoelectric properties were obtained under constant relaxation time approximation using BoltzTraP code by solving semi-classical Boltzmann transport theory [45]. A denser K-mesh of 36×36×36 was employed for the calculation of various transport properties. The Phonopy code [46] was used to ascertain the dynamical stability of all polymorphs. We were calculated the 2^{nd} and 3^{rd} order force constants using the Phono3py code [47] interfaced with the Vienna ab initio simulation package (VASP) [48] to estimate K_l. The plane wave cut-off energy was set to 560 eV, and a 9×9×9, 9×9×5, 7×7×7, 7×7×7 Gamma-centred K-point mesh was used for the *c*-, *t*-, *m*-, and *o*- HfO_2, respectively. An optimized q-mesh of 25×25×25 was used for the calculation of K_l.

3. Result and discussion:

3.1 Structural properties:

The structural parameters are optimized for all the polymorphs of HfO_2 within the parameterization of GGA by fitting Birch-Murnaghan equation of states [49], and the corresponding volume optimization curves are shown in figure 1. Our optimized lattice

Accepted Manuscript

parameters are in good agreement with the previously reported experimental results (listed in table 1) [22-24, 50]. Therefore, we have not optimized angles of the monoclinic structure but rather taken the experimental values from literature [22].

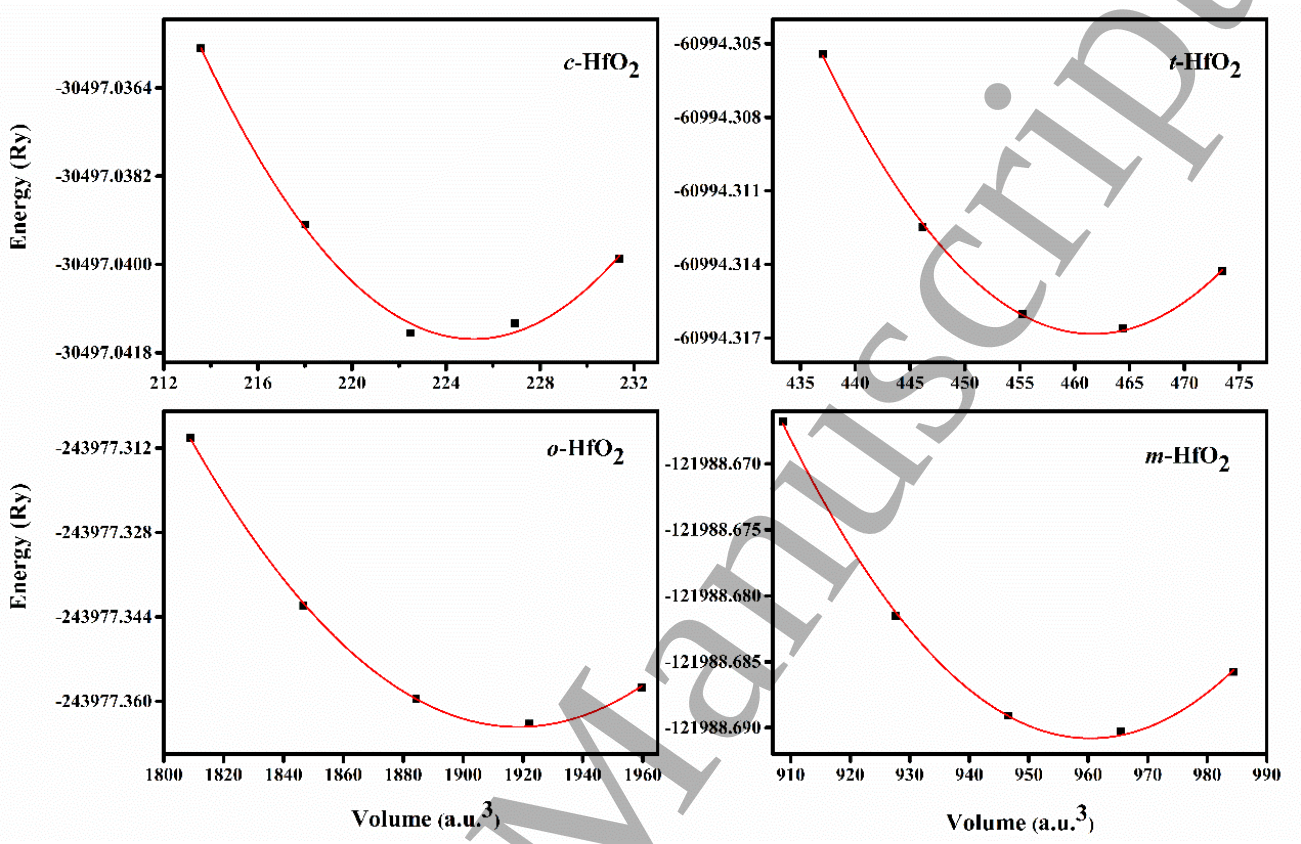


Figure 1. Energy vs Volume optimization curve of all polymorphs of HfO_2.

Table1. The optimized lattice parameters of all polymorphs of HfO_2.

Polymorphs of HfO_2	**Space group**	**Lattice parameters (in Å)**	
		Theoretical	**Experimental**
Monoclinic	$P2_1/c$	a = 5.14, b = 5.19, c = 5.33	a = 5.11, b = 5.17, c = 5.29[22]
Cubic	$Fm\bar{3}m$	a = 5.09	a = 5.09[23]
Orthorhombic	Pbca	a = 5.18, b = 10.169, c = 5.29	a = 5.107, b = 10.10, c = 5.270[50]
Tetragonal	$P4_2/nmc$	a = 3.54, c = 5.20	a = 3.62, c = 5.28[24]

Accepted Manuscript

To check the dynamical stability, we performed the phonon calculations on a 9×9×9 k-mesh using density functional perturbation theory implemented in the Phonopy code. A 2×2×2 supercell is used for monoclinic and orthorhombic structures, whereas a 3×3×3 supercell is employed for cubic and tetragonal structures to obtain force constants. The phonon band structure, as depicted in figure 2, shows no imaginary frequencies, indicating that all phases are dynamically stable.

Moreover, the mechanical stability of all the polymorphs has already been established through elastic constants [51].

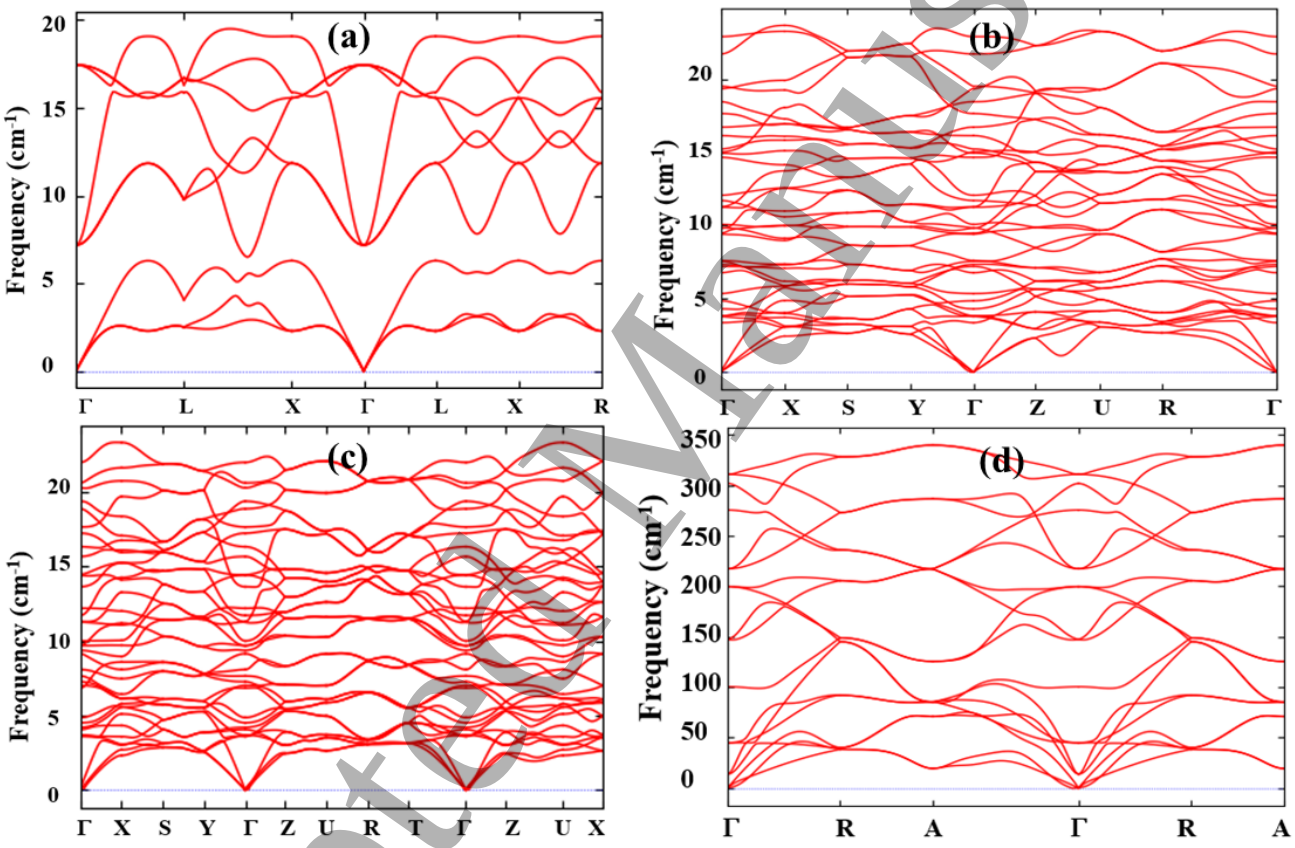


Figure 2. The phonon band structure diagram for (a) *c*-HfO_2, (b) *m*-HfO_2, (c) *o*- HfO_2, and (d) *t*- HfO_2.

3.2 Electronic properties:

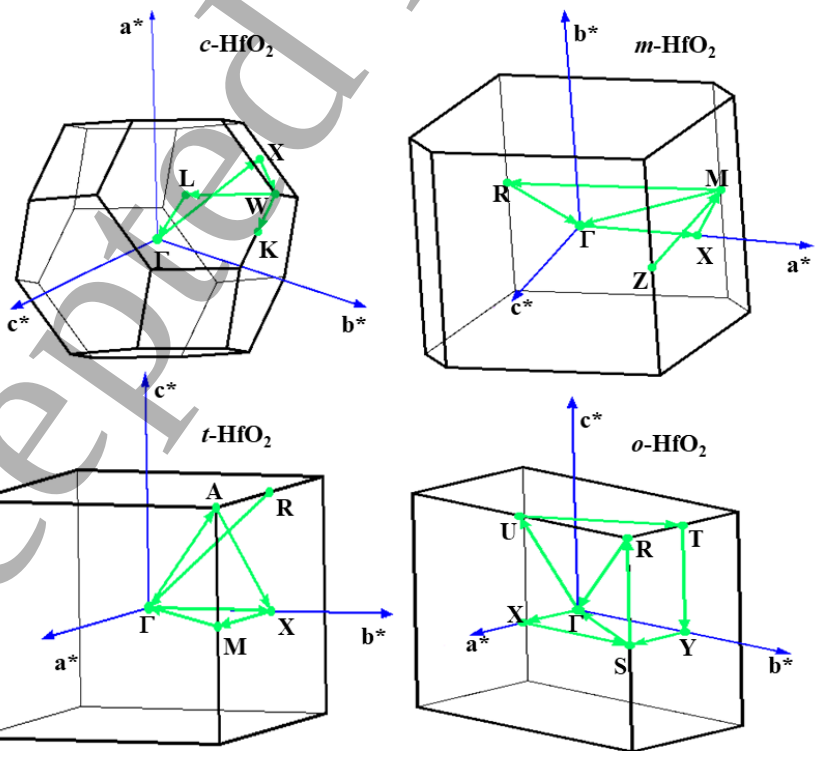
The electronic band structure calculations are essential to understanding the derived properties of polymorphic phases of HfO_2. Figure 3 shows the high symmetry points of *c*-, *m*-, *o*-, *t*- HfO_2 in the first brillioun zone that are considered in the band structure calculations. The obtained band structure for various polymorphs along the high symmetry K-path is shown in figure 4.

Accepted Manuscript

The estimated value of the band gap using GGA and TB-mBJ is listed in table 2. The calculated band gaps using TB-mBJ potential are in good agreement with the previously available experimental and theoretical results [22-23,44,52]. From the band structures, it can be seen that the top of the valance band for *m*-, *o*-, and *t*- phases are at Γ the point, and in the *c*-phase is at X point. The bottom of the conduction band is at X for *c*- and *m*- HfO_2 and at Γ for *o*- and *t*- HfO_2. The VBM of *o*-HfO_2 (inset figure 4) is shown in inset for its better appearance. Hence the phases *c*-, *o*-, and *t*-HfO_2 show the direct band gap, whilst *m*-HfO_2 exhibits an indirect band gap which has been experimentally verified in a recent study by Kumar et al. [22].

Table 2. The band gaps (with GGA and TB-mBJ approach) and effective mass for all polymorphs of HfO_2.

Polymorphs of HfO_2	**Band gaps (in eV)**				**Effective Mass**	
	Theoretical			**Experimental**	**CBM**	**VBM**
	GGA	**TB-mBJ**	**Others**			
Monoclinic	4.11	5.76	TB-mBJ 5.76[44]	5.71[22]	0.62 m_e	2.49 m_e
Cubic	3.78	5.82	TB-mBJ 5.88[44]	5.88[23]	0.73 m_e	0.28 m_e
Orthorhombic	4.46	6.30	GGA 4.30 [52]	-----	1.04 m_e	2.92 m_e
Tetragonal	4.68	6.46	TB-mBJ 6.54[44]	-----	0.59 m_e	4.85 m_e



Accepted Manuscript

Figure 3. The first brillioun zone of all polymophs of HfO_2 with high symmetry points.

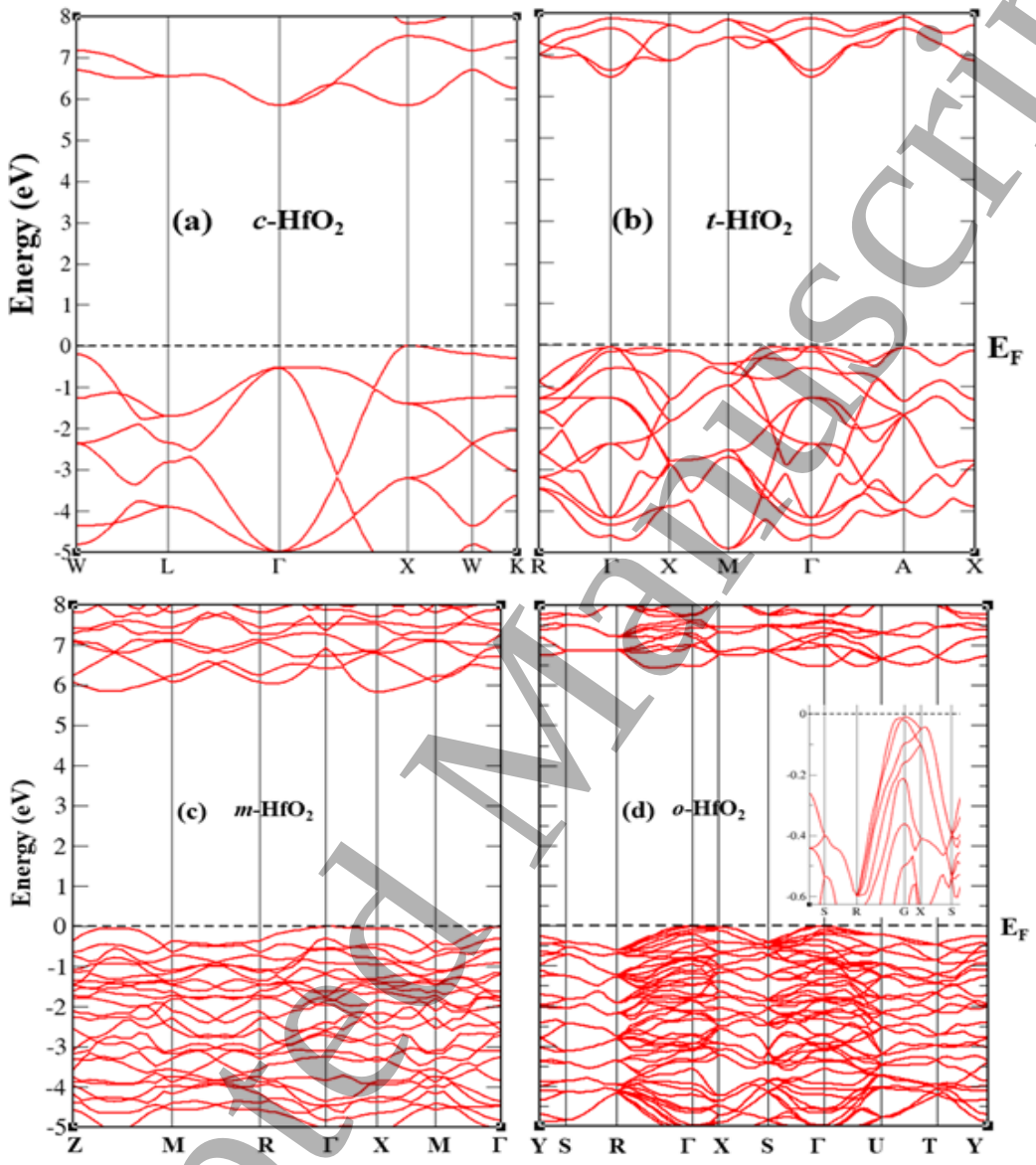


Figure 4. The band structure diagram of all the polymorphs of HfO_2. (The Fermi level (E_F) is shifted to 0 eV.) (Inset VBM of *o*-HfO_2)

For a better analysis of the band structure of all phases, we presented the partial density of states (PDOS) in figure 5. In the valance band (VB), the O-total atomic states are dominant over Hf-total atomic states and attain the maxima near the E_F. Whereas, in the conduction band (CB), Hf-total atomic states are dominant over O-atomic states in all studied energy regions. In the CB, the

Accepted Manuscript

contribution of atomic states sharply increases ~6 eV, and after that, decreases at ~8 eV in all the phases except *m*-HfO_2.

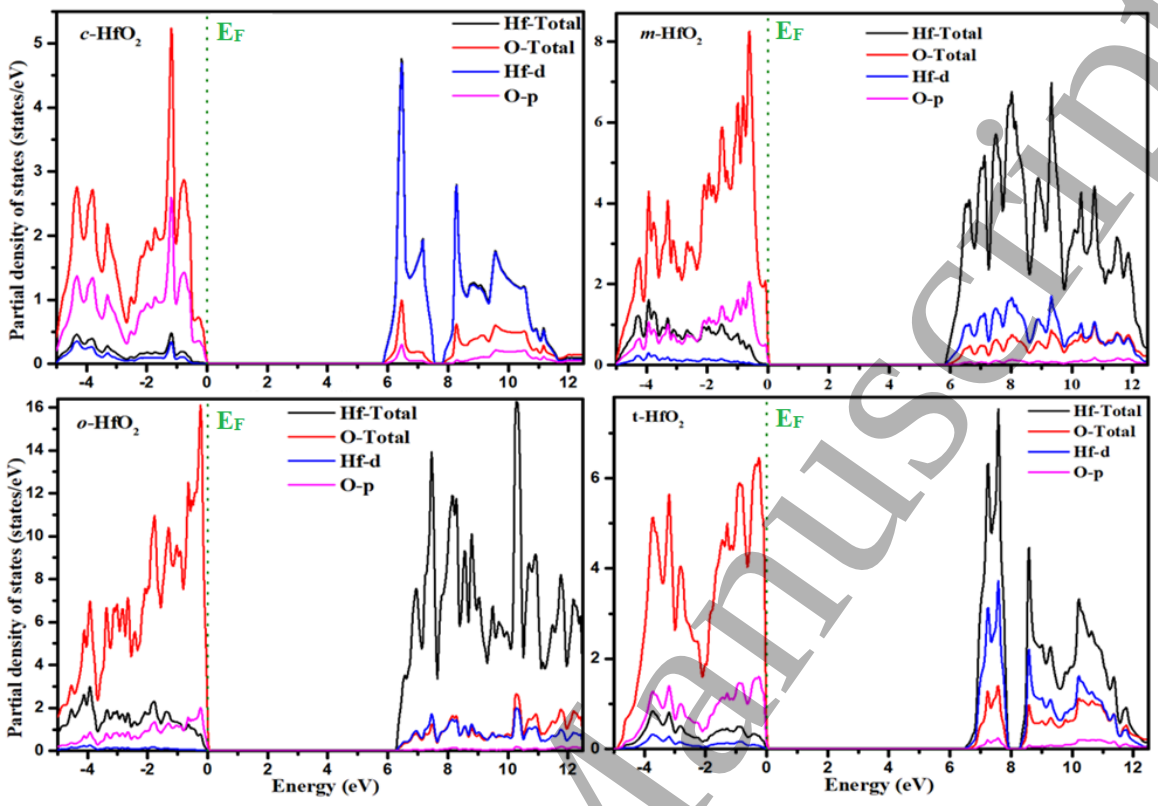


Figure 5. The partial density of states of all the polymorphs of HfO_2. (The E_F is shifted to 0 eV.)

3.3 Thermoelectric properties:

For the accurate prediction of TE properties, the precise calculation of the band gap is required [6]. As our TB-mBJ calculation give the highly accurate values of the band gap, we are highly optimistic to get a more realistic TE parameters to be compared with future experiments. The variation in the TE parameters such as *S*, σ/τ, k_e/τ, *PF* and *ZT* are analysed with the varying carrier concentration at different temperatures under the constant relaxation time approximation of 10^{-14} s. We determined the value of τ using a methodology proposed by Gudelli et al. [53] in which the experimental available value of electrical conductivity (σ) [29], and theoretical calculated value of (σ/τ) is used in the relation, $\tau = \frac{\sigma_{exp.}}{\sigma/\tau.}$. The obtained value of τ is of the order of ~10^{-14} s at room temperature. We await more experimental results for electrical conductivity at higher temperature to find the τ at elevated temperatures.

The various TE parameters are obtained using following expressions [45]:

Accepted Manuscript

$$S_{\alpha\beta}(\mu,T) = \frac{1}{eT\sigma_{\alpha\beta}(\mu,T)} \int \sigma_{\alpha\beta}(\varepsilon)(\varepsilon-\mu)\left(-\frac{\partial f(T,\varepsilon,\mu)}{\partial \varepsilon}\right) d\varepsilon \quad (1)$$

$$\kappa^{e}_{\alpha\beta}(\mu,T) = \frac{1}{e^2T\Omega} \int \sigma_{\alpha\beta}(\varepsilon)(\varepsilon-\mu)^2\left(-\frac{\partial f(T,\varepsilon,\mu)}{\partial \varepsilon}\right) d\varepsilon \quad (2)$$

$$\sigma_{\alpha\beta}(\mu,T) = \frac{1}{\Omega} \int \sigma_{\alpha\beta}(\varepsilon)\left(-\frac{\partial f(T,\varepsilon,\mu)}{\partial \varepsilon}\right) d\varepsilon \quad (3)$$

where Ω is the reciprocal space volume, f is the Fermi distribution function, e is the electron charge, ε is the carrier energy. $\sigma_{\alpha\beta}(\varepsilon)$ is conductivity tensor can be expressed as:

$$\sigma_{\alpha\beta}(\varepsilon) = \frac{1}{N}\sum_{i,k} \sigma_{\alpha\beta}(i,k)\frac{\delta(\varepsilon-\varepsilon_{i,k})}{d\varepsilon} \quad (4)$$

where N is the number of k-points in the reciprocal space.

In light of the recent experiments indicating the possibility of high carrier concentration [29-31, 33-34] in various phases of HfO_2, we analysed the effect of carrier concentration on S in the range of ~10^{18} – 10^{22} cm^{-3} for different polymorphs of HfO_2 (figure 6). The variation of S is obtained for different temperatures with carrier concentrations. As expected, it decreases monotonically with increasing carrier concentrations for all studied phases. It is also important to know the effective mass (m^*) of all the phases to validate the obtained value of S, which is given as:

$$S = \frac{8\pi^2 k_B^2}{3eh^2} m^* T \left(\frac{\pi}{3n}\right)^{\frac{2}{3}} \quad (5)$$

The m^* is evaluated by the fitting parabolic curve on the maxima of the VB and minima of the CB, which are listed in table 2. The variation in the values of m^* for different phases of HfO_2 is due to the orientation of their band curvature near the E_F, which leads to a significant contribution toward the value of S. As the m^* rises, the value of S increases. The maximum observed value of S for *c*-, *m*-, *o*-, and *t*- HfO_2 are found to be 830.81 μV/K, 867.44 μV/K, 922.62 μV/K, and 945.27 μV/K, respectively, at 300 K for 1×10^{18} cm^{-3}. We achieved a high value of S as compared to the previously reported doped (Ti, V, C, S) *c*-HfO_2 [16-19] and other oxides Cu_2O and CuO [12].

Accepted Manuscript

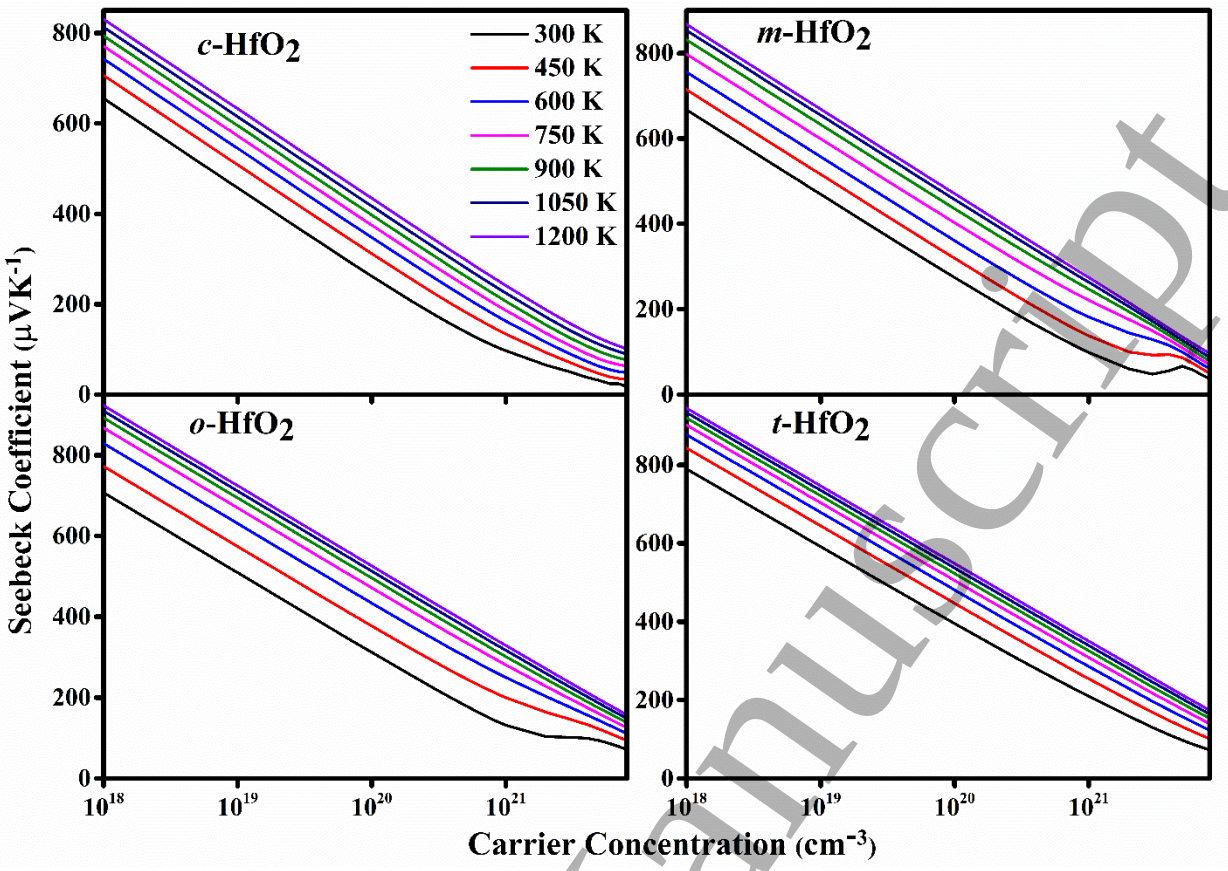


Figure 6. The variation in Seebeck Coefficient with carrier concentrations at varying temperatures for all the polymorphs of HfO_2.

The behaviour of electrical conductivity under the effect of carrier concentration at different temperatures is shown in figure 7. There is a slight variation in the values of σ/τ up to 1×10^{20} cm^{-3} after that these increases monotonically. It is noted that thermal excitation of carriers has a negligible effect on σ/τ. In the *c*-HfO_2, there is almost no effect of temperature; however, in all other phases, the value of σ/τ differ slightly at different temperatures. The highest observed value of σ/τ is 2.55×10^{20} $\Omega^{-1}m^{-1}s^{-1}$, 0.66×10^{20} $\Omega^{-1}m^{-1}s^{-1}$, 0.51×10^{20} $\Omega^{-1}m^{-1}s^{-1}$, 0.97×10^{20} $\Omega^{-1}m^{-1}s^{-1}$ at room temperature, whilst at 1200 K this slightly decreases to 2.34×10^{20} $\Omega^{-1}m^{-1}s^{-1}$, 0.54×10^{20} $\Omega^{-1}m^{-1}s^{-1}$, 0.40×10^{20} $\Omega^{-1}m^{-1}s^{-1}$, 0.85×10^{20} $\Omega^{-1}m^{-1}s^{-1}$ for *c*-, *m*-, *o*- and, *t*- HfO_2, respectively, at different carrier concentration.

Accepted Manuscript

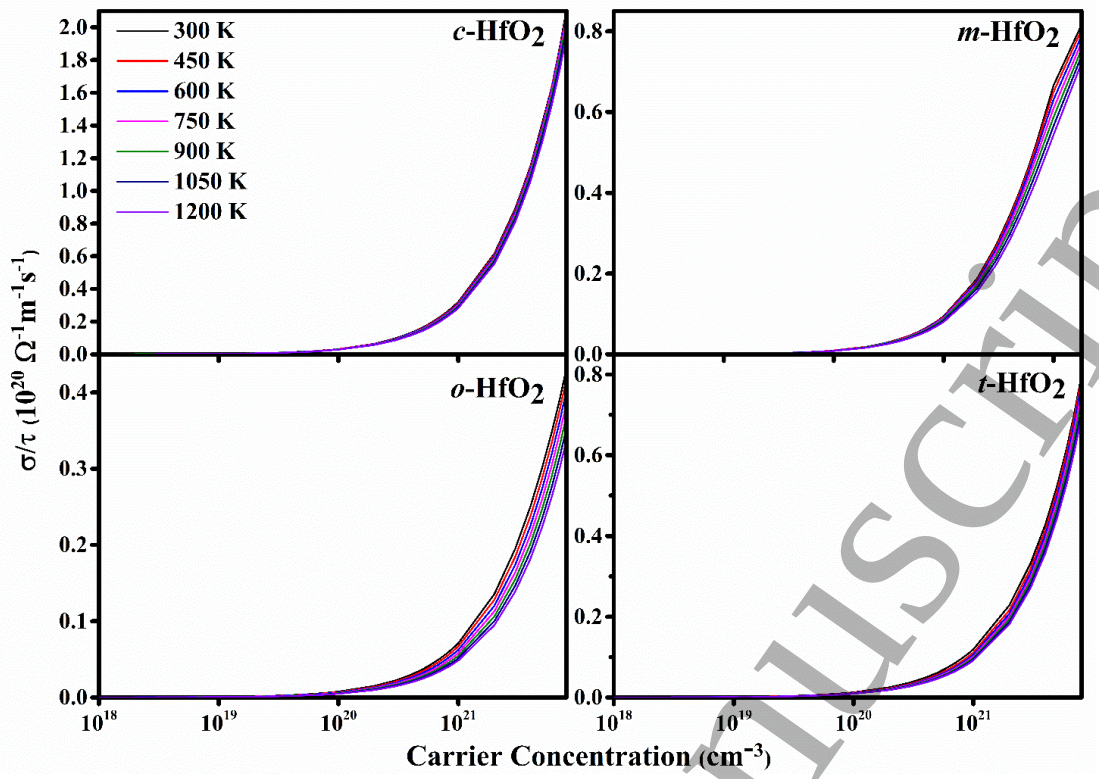


Figure 7. The variation in electrical conductivity (σ/τ) with carrier concentrations at varying temperatures for all the polymorphs of HfO_2.

The variation in electronic thermal conductivity is directly proportional to the electrical conductivity, which validates the Wiedemann Franz law, i.e., $k_e = L\sigma T$. Effect of carrier concentration at the different temperatures on k_e/τ is shown in figure 8. At room temperature, the calculated value of k_e/τ is 2.02×10^{15} W/mKs, 0.48×10^{15} W/mKs, 0.37×10^{15} W/mKs, and 0.76×10^{15} W/mKs, whereas at the higher temperature, k_e/τ increases to 8.33×10^{15} W/mKs, 1.49×10^{15} W/mKs, 1.40×10^{15} W/mKs, and 3.95×10^{15} W/mKs at 1200 K, for *c*-, *m*-, *o*- and, *t*-HfO_2, respectively at carrier concentration. The obtained values of σ/τ are significantly increased as compared to Cu_2O, NiO, and CuO [12].

Accepted Manuscript

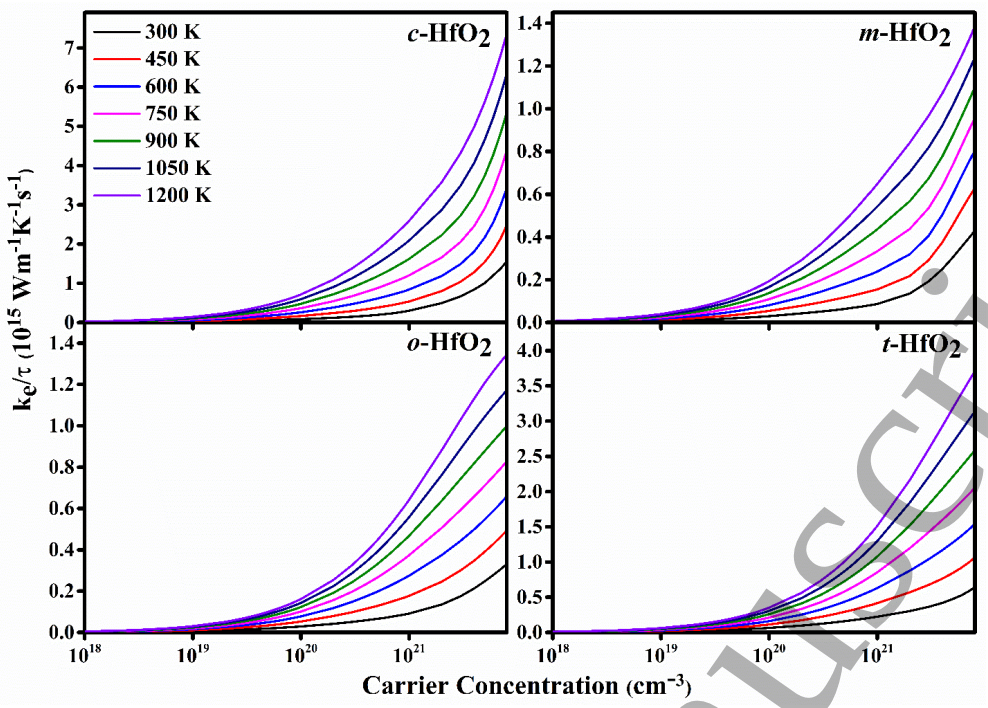


Figure 8. The variation in electronic thermal conductivity (k_e/τ) with carrier concentrations at varying temperatures for all the polymorphs of HfO_2.

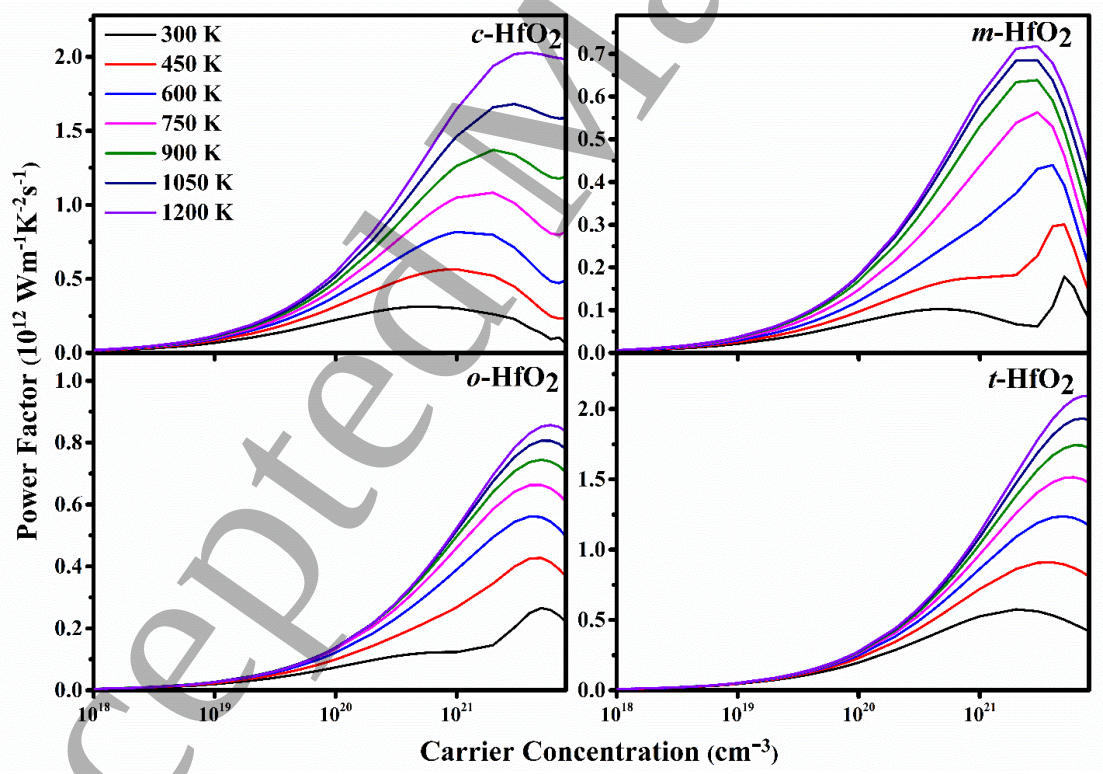


Figure 9. The variation in Power factor with carrier concentrations at varying temperatures for all the polymorphs of HfO_2.

Accepted Manuscript

An increase in the value of *PF* is observed with the increasing carrier concentration, as shown in figure 9. The calculated value of *PF* is purely contributed by the *S* and *s*. As we move towards the higher temperature, the increment in *PF* is observed; the highest estimated value of *PF* is 2.020×10^{12} W/mK^{2}s at 4×10^{21} cm^{-3}, 0.71×10^{12} W/mK^{2}s at 3×10^{21} cm^{-3}, 0.85×10^{12} W/mK^{2}s at 6×10^{21} cm^{-3}, and 2.09×10^{12} W/mK^{2}s at 8×10^{21} cm^{-3}, for *c*-, *m*-, *o*- and, *t*- HfO_2, respectively. Therefore, it can be seen that the optimized carrier concentration for the *PF* is of the order of 10^{21} cm^{-3}.

The lattice part of the thermal conductivity plays a major role in maximizing the TE performance as it can be tuned independently, unlike *S*, σ/τ, k_e/τ, which are interdependent on each other. The K_l decreases with the increase in temperature, as shown in figure 10. The room temperature K_l for *c*-, *m*-, *o*-, and, *t*- HfO_2 are obtained as 5.56 W/mK, 2.87 W/mK, 4.32 W/mK, and 1.75 W/mK, respectively, and this decreases to 1.58 W/mK, 0.92 W/mK, 1.12 W/mK, 0.53 W/mK, respectively at 1200 K. The high value of K_l of the *c*-HfO_2 phase as compared to others may be due to its isotopic nature and less overlapping between its acoustic and optical modes. A significant extent of overlapping in *m*-, *o*-, and *t*-HfO_2 is observed as visible in the phonon band structure diagram. The low value of K_l indicates that HfO_2 polymorphs can be promising TE materials at elevated temperature.

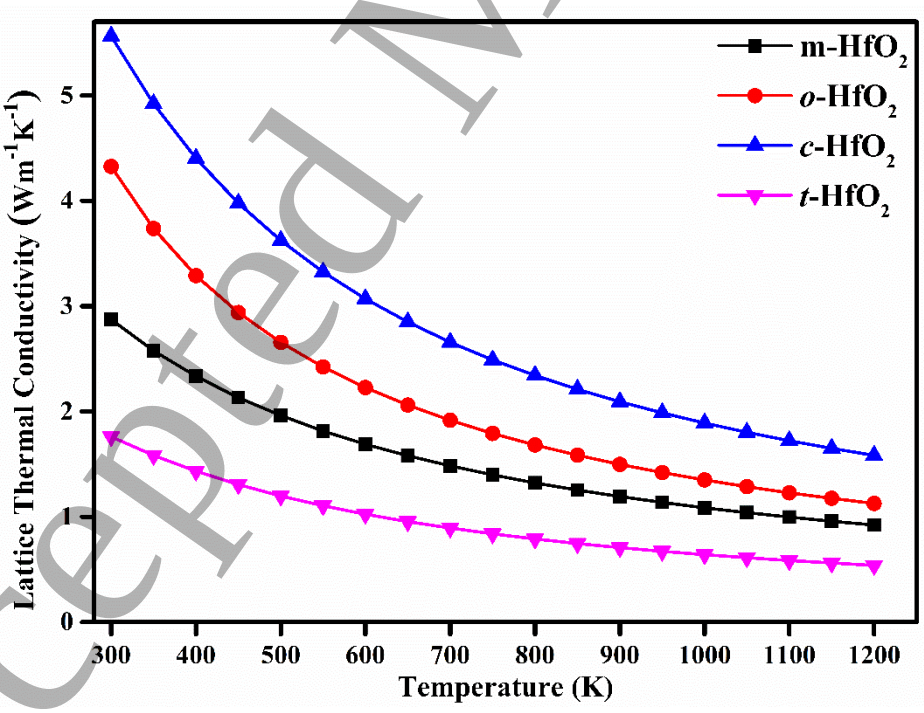


Figure 10. The behaviour of lattice thermal conductivity with varying temperatures for all the polymorphs of HfO_2.

Accepted Manuscript

Based upon the values of *PF* and K_{total}, the carrier concentration dependent *ZT* is determined for polymorphs at different fixed temperatures, as shown in figure 11. As the temperature rises, the *ZT* rises with it. It is observed that the *ZT* increases with an increase in carrier concentration up to 10^{21} cm^{-3} and after that decreases. The peak value of *ZT* is 0.77 at 3×10^{20} cm^{-3}, 0.97 at 1×10^{21} cm^{-3}, 0.83 at 2×10^{21} cm^{-3}, 0.87 at 5×10^{21} cm^{-3} at the 1200 K for *c*-, *m*-, *o*- and *t*- HfO_2, respectively. The highest value of *ZT* of the *m*-HfO_2 phase is mainly governed by its low K_{total} among all polymorphs.

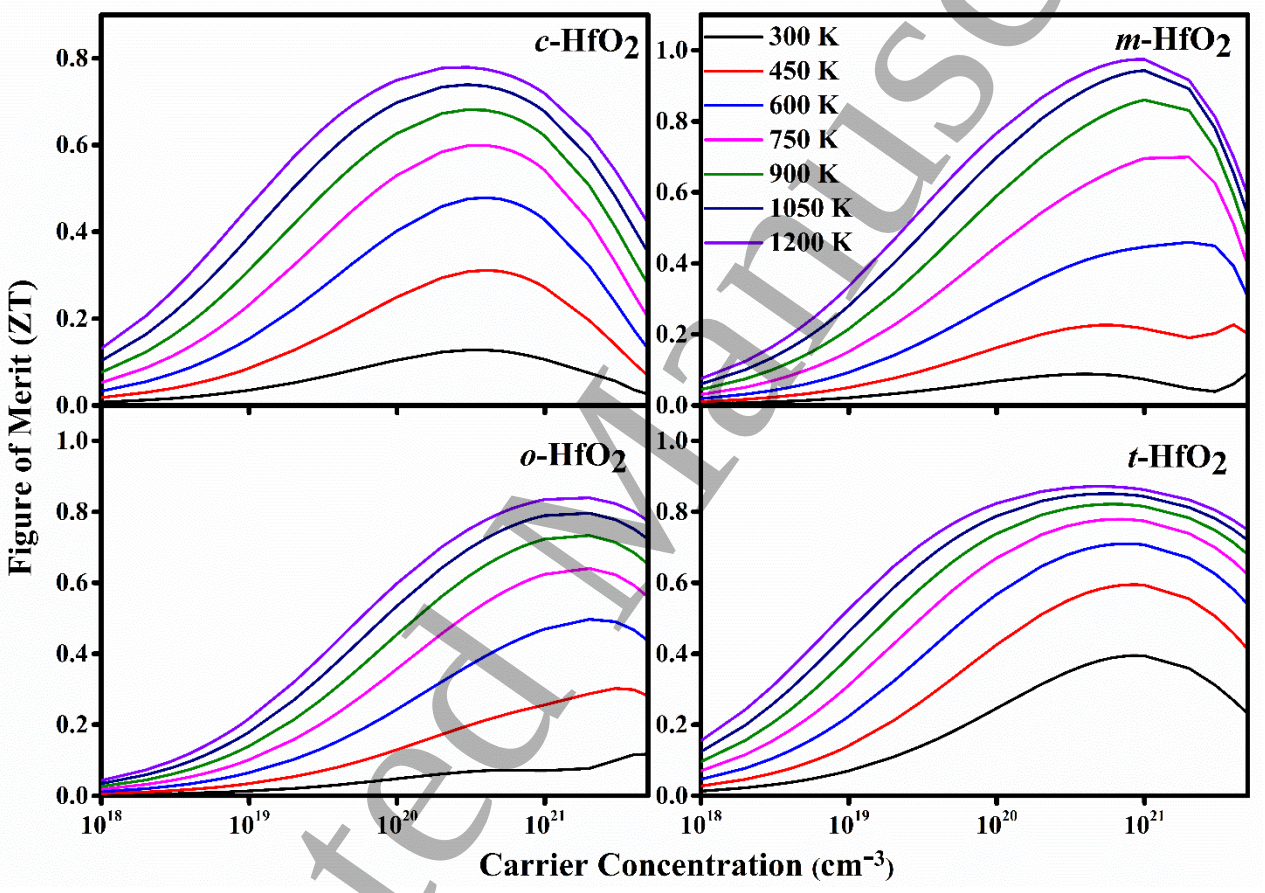


Figure 11. The variation in the figure of merit with carrier concentrations at varying temperatures for all the polymorphs of HfO_2.

After obtaining the optimized carrier concentrations (3×10^{20} cm^{-3}, 1×10^{21} cm^{-3}, 2×10^{21} cm^{-3}, 5×10^{20} cm^{-3} for *c*-, *m*-, *o*-, and *t*- HfO_2, respectively), we investigated the effect of temperature on ZT at these optimized carrier concentrations (figure 12). It is observed that *ZT* increases with increasing temperature. At room temperature, the calculated values of *ZT* are 0.12, 0.07, 0.07,0.39 and a sharp increment is observed as we move towards high temperatures; at 1200 K, these values increase to 0.77, 0.97, 0.83, 0.87 for *c*-, *m*-, *o*- and, *t*- HfO_2, respectively. At

Accepted Manuscript

750 K, the reported value of *ZT* in *m*- and *t*-HfO_2 is nearly equal. It is to noted that we got high figure of merit at the optimized carrier concentration in HfO_2 as compare to BiCuSeO (*ZT*~0.70) [54], Ag-doped $Ca_3Co_4O_9$ (*ZT* = 0.50) [55], La-doped $SrTiO_3$ (*ZT* = 0.27) [56], Al-doped ZnO (*ZT* = 0.44) [57], Al and Ga co-doped ZnO (*ZT* = 0.65) [58]. The various studied polymorphs of HfO_2 found suitable for high temperature TE energy harvesting applications. The carrier concentration mediated high *ZT* for HfO_2 is experimentally viable due to its high temperature stability [1200 ^oC – 2500 ^oC] and an attainable carrier concentration (~10^{21} cm^{-3}) with naturally occurring oxygen defects [29-31, 33-34].

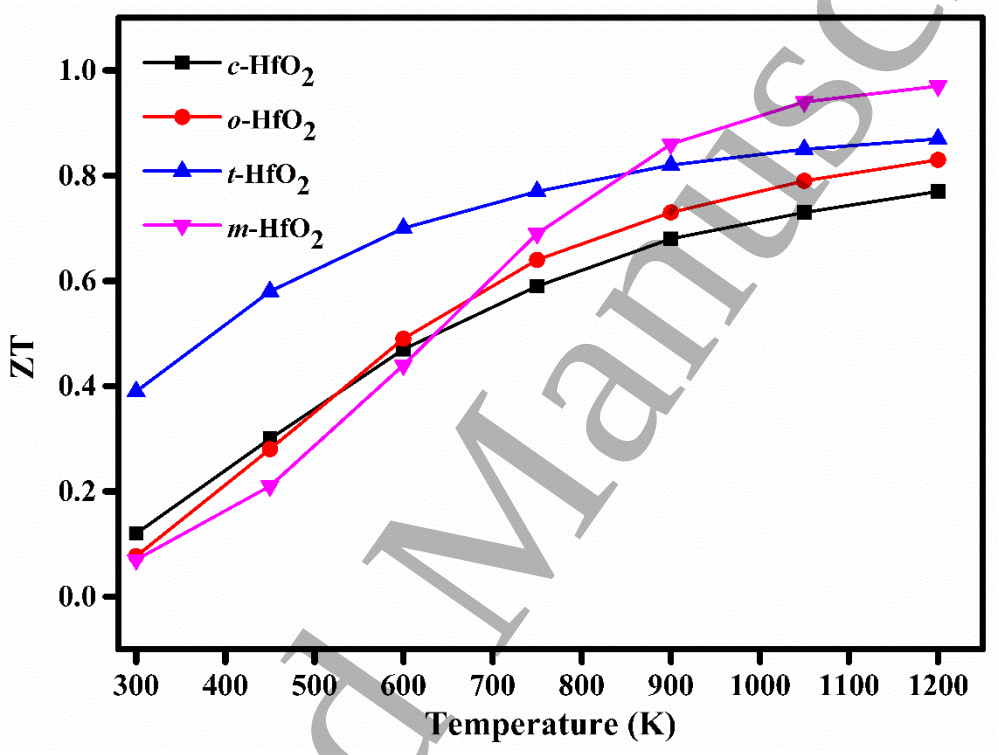


Figure 12. The variation in figure of merit with temperature at the optimized carrier concentrations (3×10^{20} cm^{-3}, 1×10^{21} cm^{-3}, 2×10^{21} cm^{-3}, 5×10^{20} cm^{-3} for *c*-, *m*-, *o*- and, *t*- phases, respectively).

4. Conclusion:

We have explored, using Density Functional Theory, the structural, electronic, phonon dynamics, and transport properties of HfO_2 polymorphs, cubic (*c*), monoclinic (*m*), tetragonal (*t*) and orthorhombic (*o*), for their possible application in high temperature thermoelectric energy harvesting. This study is inspired by few recent experimental reports which have established the possibility of realizing high carrier concentration in some of these polymorphs.

Accepted Manuscript

The dynamical stability of all the polymorphs has been assessed by phonon dispersion calculations. The electronic band structure and transport properties have calculated using the highly accurate TB-mBJ potentials for electron-electron exchange and correlation. Our calculations have revealed that the *c*-, *o*-, and *t*- phases have a direct band gap, and the *m*-phase has an indirect band gap. These findings are in good agreement with the available experimental results. The positive values of the Seebeck coefficient for all the HfO_2 polymorphs at various temperatures ensure their *p*-type behaviour. Among all the studied polymorphs, the *t*-HfO_2 shows the lowest value of lattice thermal conductivity 1.75 W/mK at 300 K which reduces to 0.53 W/mK at 1200 K. Owing to large power factor and low lattice thermal conductivity, the high figure of merit, i.e., 0.97, 0.87, 0.83, and 0.77 has observed at 1200 K for the optimized carrier concentrations 1×10^{21} cm^{-3}, 2×10^{21} cm^{-3}, 5×10^{20} cm^{-3}, and 3×10^{20} cm^{-3} for the *m*-, *o*-, *t*-, and, *c*-polymorphs of HfO_2, respectively. These values of optimized carrier concentration are in close agreement with the available experimental data. Our results have demonstrated that the TE parameters can be manipulated in a favourable way at selected carrier concentration and temperature to get enhanced TE performance at higher temperature.

Acknowledgement:

We acknowledge National Supercomputing Mission (NSM) for providing computing resources of 'PARAM Seva' at IIT Hyderabad, which is implemented by C-DAC and supported by the Ministry of Electronics and Information Technology (MeitY) and Department of Science and Technology (DST), Government of India.

References:

[1] Qazi A, Hussain F, Rahim N A B D, Hardaker G, Alghazzawi D, Shaban K, Haruna K 2019 Towards sustainable energy: a systematic review of renewable energy sources, technologies, and public opinions IEEE Access 7 63837e63851 https://doi.org/10.1109/ACCESS.2019.2906402.

[2] Elavarasan R M, Shafiullah G, Padmanaban S, Kumar N M, Annam A, Vetrichelvan A M, Popa L M, Nielsen J B H 2020 A comprehensive review on renewable energy development, challenges, and policies of leading Indian states with an international perspective IEEE Access 8 74432e74457 https://doi.org/10.1109/ACCESS.2020.2988011.

[3] Østergaard P A, Duic N, Noorollahi Y, Mikulcic H, Kalogirou S, 2020 Sustainable development using renewable energy technology Renew. Energy 146 2430e2437 https://doi.org/10.1016/j.renene.2019.08.09.

[4] Enescu D 2019 Thermoelectric energy harvesting: basic principles and applications Green Energy Adv. https://doi.org/10.5772/intechopen.83495.

Accepted Manuscript

[5] Sharma P K, Senguttuvan T D, Sharma V K, Chaudhary S 2021 Revisiting the thermoelectric properties of lead telluride Mater. Today Energy 21 100713 https://doi.org/10.1016/j.mtener.2021.100713.
[6] Sun Y, Liu Y, Li R, Li Y, Bai S 2022 Strategies to Improve the Thermoelectric Figure of Merit in Thermoelectric Functional Materials Front. Chem. 10 865281 doi: 10.3389/fchem.2022.865281
[7] Zhu Y, Wang D, Hong T, Hu L, Ina T, Zhan S, Qin B, Shi H, Su L, Gao L, Zhao L D 2022 Multiple valence bands convergence and Strong phonon scattering lead to high thermoelectric performance in p-type PbSe Nat. Commun. 13 4179 doi: 10.1038/s41467-022-31939-4
[8] Chen Z, Zhang X, Pei Y 2018 Manipulation of Phonon Transport in Thermoelectrics Adv. Materials 30 1705617 https://doi.org/10.1002/adma.201705617.
[9] Witting I T, Chasapis T C, Ricci F, Peters M, Heinz N A, Hautier G, Snyder G J 2019 The Thermoelectric Properties of Bismuth Telluride Adv. Electron. Mater. 5 1800904 https://doi.org/10.1002/aelm.201800904
[10] Xiao Y, Zhao L D 2018 Charge and phonon transport in PbTe-based thermoelectric materials npj Quantum Mater. 3 55 https://doi.org/10.1038/s41535-018-0127-y
[11] Yin Y, Tudu B, Tiwari A 2017 Recent advances in oxide thermoelectric Vacuum 146 356-374 https://doi.org/10.1016/j.vacuum.2017.04.015
[12] Linnera J, Sansone G, Maschio L, Karttunen A J 2018 Thermoelectric Properties of p-Type Cu_2O, CuO, and NiO from Hybrid Density Functional Theory J. Phys. Chem. C 122 15180–15189 https://doi.org/10.1021/acs.jpcc.8b04281
[13] Peng L, Miao N, Wang G, Zhou J, Elliott S R, Sun Z 2021 Novel metal oxides with promising high-temperature thermoelectric performance J. Mater. Chem. C 9 12884-12894 https://doi.org/10.1039/D1TC02404C
[14] Akram R, Khan J, Rafique S, Hussain M, Maqsood A, Naz A A 2021 Enhanced thermoelectric properties of single phase Na doped Na_XCoO_2 thermoelectric material Mater. Lett. 300 130180 https://doi.org/10.1016/j.matlet.2021.130180
[15] Pashkevich A V, Fedotov A K, Poddenezhny E N, Bliznyuk L A, Fedotov J A, Basov N A, Kharchanka A A, Zukowski P, Koltunowicz T N, Korolik O V, Fedotova V V 2022 Structure, electric and thermoelectric properties of binary ZnO-based ceramics doped with Fe and Co J. Alloys Compd. 895 162621 https://doi.org/10.1016/j.jallcom.2021.162621
[16] Cao Z, Fan T, Hou X, Niu J, Sharma R, Dar S A 2019 Evaluation of structural, electronic, thermoelectric, and optical results of C-doped HfO_2 by first-principle's investigation Int. J. Energy Res. 44 1–14 https://doi.org/10.1002/er.4878
[17] Sharma R, Dar S A, Srivastava V 2021 Half-metallic ferromagnetism and thermoelectric properties of vanadium doped Hf1−xVxO2(x = 0, 0.25, 0.50, 0.75) alloys by first principles perspective Phys. Lett. A 411 127559 https://doi.org/10.1016/j.physleta.2021.127559
[18] Kumar R, Vij A, M. Singh M 2022 Electronic, thermoelectric, and optical studies of cubic $Hf1\text{-}xTixO_2$: An attempt to enhance the key parameters 307 122829 https://doi.org/10.1016/j.jssc.2021.122829
[19] Kumar R, Kumar R, Vij A, Singh M 2022 A first-principle study of electronic, thermoelectric, and optical properties of sulfur doped c-HfO2 Phys. Scr. 97 075813 https://doi.org/10.1088/1402-4896/ac7678

Accepted Manuscript

[20] Baik M, Kang H K, Kang Y S, Jeong K S, An Y, Choi S, Kim H, Song J D, Cho M H 2017 Electrical properties and thermal stability in stack structure of HfO2/Al2O3/InSb by atomic layer deposition Sci Rep. 7 11337 https://doi.org/10.1038/s41598-017-09623-1

[21] Leiming H L, Guan F S, Peng F, Zhang Z, Chen H, Zhang W, Lu C 2020 Insights into the Bond Behavior and Mechanical Properties of Hafnium Carbide under High Pressure and High Temperature Inorg. Chem. 60 515–524 https://doi.org/10.1021/acs.inorgchem.0c02800

[22] Kumar R, Vij A, Singh M 2021 Defects assisted luminescence in *m*-HfO_2 nanocrystals: An experimental and theoretical study Optik 248 168121 https://doi.org/10.1016/j.ijleo.2021.168121

[23] Kumar N, George B P A, Abrahamse H, Parashar V, Ray S S, Ngila J C 2017 A novel approach to low-temperature synthesis of cubic HfO_2 nanostructures and their cytotoxicity Sci Rep 7 9351 https://doi.org/10.1038/s41598-017-07753-0

[24] Tobase T, Yoshiasa A, Arima H, Sugiyama K, Ohtaka O, Nakatani T, Funakoshi K, Kohara S 2018 Pre-Transitional Behavior in Tetragonal to Cubic Phase Transition in HfO2 Revealed by High Temperature Diffraction Experiments Phys. Status Solidi B 255 1800090 https://doi.org/10.1002/pssb.201800090

[25] Banerjee D, Sewak R, Dey C C, Toprek D, Pujari P K 2021Orthorhombic phases in bulk pure HfO_2: Experimental observation from perturbed angular correlation spectroscopy Mater. Today Commun. 26 101827 https://doi.org/10.1016/j.mtcomm.2020.101827

[26] Pathak S, Mandal G, Das P, Dey A B 2020 Structural characteristics of HfO_2 under extreme conditions, Mater. Chem. Phys. 255 123633 https://doi.org/10.1016/j.matchemphys.2020.123633

[27] Xu X, Huang F T, Qi Y 2021 Kinetically stabilized ferroelectricity in bulk single-crystalline HfO2:Y Nat. Mater. 20 826–832 https://doi.org/10.1038/s41563-020-00897-x

[28] Villa I, Moretti F, Fasoli M, Rossi A, Hattendorf B, Dujardin C, Niederberger M, Vedda A, Lauria A, 2019 The Bright X-Ray Stimulated Luminescence of HfO_2 Nanocrystals Activated by Ti Ions Adv. Optical Mater. 8 1901348 https://doi.org/10.1002/adom.201901348

[29] Hildebrandt E, Kurian J, Mueller M M, Schroeder T, Kleebe H J, Lambert A 2011 Controlled oxygen vacancy induced p-type conductivity in HfO{2-x} thin films Appl. Phys. Lett. 99 112902 https://doi.org/10.1063/1.3637603

[30] Kaiser N, Vogel T, Zintler A, Petzold S, Arzumanov A, Piros E, Eilhardt R, Luna L M, Alff L 2022 Defect-Stabilized Substoichiometric Polymorphs of Hafnium Oxide with Semiconducting Properties ACS Appl. Mater. Interfaces 14 1290-1303 https://doi.org/10.1021/acsami.1c09451

[31] Hadacek N, Nosov A, Ranno L, Strobel P, Galéra R M 2007 Magnetic Properties of Hfo2 Thin Films J. Phys.: Condens. Matter 19 486206 https://doi.org/10.1088/0953-8984/19/48/486206

[32] González L A, Barboza S G, Lujano E V, Galicia J L R, Cerda L A G 2020 Mn-modified HfO_2 nanoparticles with enhanced photocatalytic activity Ceram. Int. 46 13466-13473 https://doi.org/10.1016/j.ceramint.2020.02.130.

Accepted Manuscript

[33] Perevalov T V, Prosvirin I P, Suprun E A, Mehmood F, Mikolajick T, Schroeder U, Gritesenko V A 2021 The atomic and electronic structure of $Hf_{0.5}Zr_{0.5}O_2$ and $Hf_{0.5}Zr_{0.5}O_2$:La films J. Sci.: Adv. Mater. Devices 6 595-600 https://doi.org/10.1016/j.jsamd.2021.08.001

[34] Islamov D R, Perevalov T V 2019 Effect of oxygen vacancies on the ferroelectric $Hf_{0.5}Zr_{0.5}O_2$ stabilization: DFT simulation Microelectron. Eng. 216 111041 https://doi.org/10.1016/j.mee.2019.111041

[35] Duana B, Lia Y, Lia J, Gaoa Y, Zhaia P, Yangc J, Lua Z, Yanga H, Wanga H, Li G, 2020 Regulation of oxygen vacancy and reduction of lattice thermal conductivity in ZnO ceramic by high temperature and high pressure method Ceram. Init. 46 26176-26181 https://doi.org/10.1016/j.ceramint.2020.07.115

[36] Ren G K, Lan J L, Ventura K J, Tan X, Lin Y H, Nan C W 2016 Contribution of point defects and nano-grains to thermal transport behaviours of oxide-based thermoelectrics npj Comput. Mater. 2 16023 https://doi.org/10.1038/npjcompumats.2016.23

[37] Chen L, Zhang Y, Wang X, Jalan B, Chen S, Hou Y 2018 Roles of Point Defects in Thermal Transport in Perovskite Barium Stannate J. Phys. Chem. C 122 11482−11490 https://doi.org/10.1021/acs.jpcc.8b00653

[38] Scott E A, Gaskins J T, King S W, Hopkins P E 2018 Thermal conductivity and thermal boundary resistance of atomic layer deposited high-k dielectric aluminum oxide, hafnium oxide, and titanium oxide thin films on silicon APL Mater. 6 058302 https://doi.org/10.1063/1.5021044

[39] Panzer M A, Shandalov M, Rowlette J A, Oshima Y, Chen Y W, McIntyre P C, Goodson K E 2009 Thermal Properties of Ultrathin Hafnium Oxide Gate Dielectric Films, IEEE Electron Device Lett. 30 1269-1271 https://doi.org/10.1109/led.2009.2032937

[40] Blaha P, Schwarz K, Madsen G K H, Kvasnicka D, Luitz J 2018 wien2k, an Augment Plane Wave+ Local Orbitals Program for Calculating Crystal Properties, User's Guide Vienna Austria ISBN 3-9501031-1-2

[41] Perdew J P, Burke K, Ernzerhof M 1996 Generalized Gradient Approximation Made Simple Phys Rev Lett. 77 3865-3868 https://doi.org/10.1103/PhysRevLett.78.1396

[42] Borlido P, Schmidt J, Huran A W, Tran F, A. L. Marques M A L, Botti S 2020 Exchange-correlation functionals for band gaps of solids: benchmark, reparametrization and machine learning. npj Comput. Mater. 6 96 https://doi.org/10.1038/s41524-020-00360-0

[43] Tran F, Blaha P 2009 Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential Phys Rev Lett. 102 226401 https://doi.org/10.1103/PhysRevLett.102.226401

[44] Ondracka P, Holec D, Nečas D, Zajíčková L 2015 Accurate prediction of band gaps and optical properties of HfO2. J. Phys. D: Appl. Phys. 49 305301 https://doi.org/10.1088/0022-3727/49/39/395301.

[45] Madsen G K, Singh D J 2006 BoltzTrap. A code for calculating band-structure dependent qualities Comput. Phys. Commun. 175 67- 71 https://doi.org/10.1016/j.cpc.2006.03.007

[46] Togo A, Tanaka I 2015 First principles phonon calculations in materials science, Scr. Mater. 108 1-5 (2015) http://dx.doi.org/10.1016/j.scriptamat.2015.07.021

Accepted Manuscript

[47] Togo A, Chaput L, Tanaka I, 2015 Distributions of phonon lifetimes in Brillouin zones Phys. Rev. B 91 094306 https://doi.org/10.1103/PhysRevB.91.094306
[48] Kresse G, D. Joubert D 1999 From ultrasoft pseudopotentials to the projector augmented-wave method Phys. Rev. B 59 1758 https://doi.org/10.1103/PhysRevB.59.1758
[49] Birch F (1947) Finite Elastic Strain of Cubic Crystals Phys. Rev. 71, 809 https://doi.org/10.1103/PhysRev.71.809
[50] Banerjee D, Sewak R, Dey C C, Toprek D, Pujari P K 2021Orthorhombic phases in bulk pure HfO2: Experimental observation from perturbed angular correlation spectroscopy Mater. Today Commun. 26 101827 https://doi.org/10.1016/j.mtcomm.2020.101827
[51] Wu R, Zhou B, Li Q, Jiang Z Y, Weng W B, Ma W Y, Zhang X D 2012 Elastic and vibrational properties of monoclinic HfO2 from first-principles study J. Phys. D: Appl. Phys. 45 125304 https://doi.org/10.1088/0022-3727/45/12/125304.
[52] Jaffe J E, Bachorz R A, Gutowski M 2005 Low-temperature polymorphs of ZrO2 and HfO2: A density-functional theory study Phys. Rev. B 72 144107 https://doi.org/10.1103/PhysRevB.72.144107
[53] Gudelli V K, Kanchana V, Vaitheeswaran G, Svane A, Christensen N E 2013 Thermoelectric properties of chalcopyrite type $CuGaTe_2$ and chalcostibite $CuSbS_2$ J. Appl. Phys 114 223707 https://doi.org/10.1063/1.4842095
[54] Li F, Li J F, Zhao L D, Xiang K, Liu Y, Zhang B P, Lin Y H, Nan C W, Zhu H M 2012 Polycrystalline BiCuSeO oxide as a potential thermoelectric material Energy Environ. Sci. 5 7188-7195 https://doi.org/10.1039/C2EE21274A
[55] Wang Y, Sui Y, Cheng J G, Su W H 2009 Comparison of the high temperature thermoelectric properties for Ag-doped and Ag-added Ca3Co4O9 J. Alloys Compd. 477 817–821 https://doi.org/10.1016/j.jallcom.2008.10.162
[56] Ohta S, Nomura T, Ohta H, Koumoto K 2005 High-temperature carrier transport and thermoelectric properties of heavily La- or Nb-doped SrTiO3 single crystals, J. Appl. Phys. 97 034106 https://doi.org/10.1063/1.1847723
[57] Jood P, Mehta R J, Zhang Y L, Peleckis G, Wang X L, Siegel R W, Tasciuc T B, Dou S X, Ramanath G 2011 Al-Doped Zinc Oxide Nanocomposites with Enhanced Thermoelectric Properties Nano Lett. 11 4337–4342 https://doi.org/10.1021/nl202439h
[58] Ohtaki M, Araki K, Yamamoto K 2009 High thermoelectric performance of dually doped ZnO ceramics J. Electron. Mater. 38 1234–1238 https://doi.org/10.1007/s11664-009-0816-1

Accepted Manuscript

arXiv:2210.01435v1 [math.CV] 4 Oct 2022

CERTAIN COEFFICIENT PROBLEMS OF $\mathcal{S}_e^*$ AND $\mathcal{C}_e$

S. SIVAPRASAD KUMAR AND NEHA VERMA

Abstract. In the present investigation, we consider the following classes of starlike and convex functions associated with an exponential function, respectively given by

$$\mathcal{S}_e^* = \left\{ f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec e^z \right\} \quad \text{and} \quad \mathcal{C}_e = \left\{ f \in \mathcal{A} : 1 + \frac{zf''(z)}{f'(z)} \prec e^z \right\},$$

to establish certain coefficient related problems such as estimation of sharp bound of third Hankel determinant, bounds of sixth and seventh coefficients as well as the fourth order Hankel determinant.

1. Introduction

Let $\mathcal{A}$ be the class of normalized analytic functions defined on the open unit disk $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$, given by

$$f(z) = z + \sum_{n=2}^{\infty} a_n z^n \tag{1.1}$$

and let $\mathcal{S}$ be the subclass of $\mathcal{A}$ consisting of univalent functions. Let $\mathcal{P}$ be the class of analytic functions defined on $\mathbb{D}$ with positive real part, given by $p(z) = 1 + \sum_{n=1}^{\infty} p_n z^n$. Let h and g be two analytic functions. We say h_1 is subordinate to h_2, denoted by $h_1 \prec h_2$, if there exists a Schwarz function w with $w(0) = 0$ and $|w(z)| \leq |z|$ such that $h_1(z) = h_2(w(z))$. Bieberbach conjecture [3][Page no.] contributed significantly to the growth of geometric function theory and the rise of new coefficient problems, we refer [3] for more information. Since then, more number of subclasses of $\mathcal{S}$ involving $\mathcal{S}^*$ and $\mathcal{C}$ of starlike and convex functions respectively have been introduced. In 1992, Ma and Minda [14] introduced the following two classes:

$$\mathcal{S}^*(\varphi) = \left\{ f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec \varphi(z) \right\} \tag{1.2}$$

and

$$\mathcal{C}(\varphi) = \left\{ f \in \mathcal{A} : 1 + \frac{zf''(z)}{f'(z)} \prec \varphi(z) \right\}, \tag{1.3}$$

which unifies various subclasses of $\mathcal{S}^*$ and $\mathcal{C}$ respectively. Here φ is an analytic univalent function satisfying the conditions $\operatorname{Re} \varphi(z) > 0$, $\varphi(\mathbb{D})$ symmetric about the real axis and starlike with respect to $\varphi(0) = 1$ with $\varphi'(0) > 0$.

The concept of Hankel determinants was introduced in [17] and is a topic of interest for many authors even today. The definition of the q^{th} Hankel determinants $H_{q,n}(f)$ of analytic functions $f \in \mathcal{A}$ of the form (1.1) under the assumption that $a_1 := 1$ is as follows:

$$H_{q,n}(f) = \begin{vmatrix} a_n & a_{n+1} & \dots & a_{n+q-1} \\ a_{n+1} & a_{n+2} & \dots & a_{n+q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n+q-1} & a_{n+q} & \dots & a_{n+2q-2} \end{vmatrix} \tag{1.4}$$

2010 *Mathematics Subject Classification.* 30C45, 30C50.

Key words and phrases. Coefficient estimate, Exponential, Starlike, Hankel determinants .

for $n, q \in \mathbb{N}$. The third order Hankel determinant expression, which is denoted by $H_{3,1}(f)$, is derived from (1.4) by taking $q = 3$ and $n = 1$, defined as:

$$H_{3,1}(f) = 2a_2a_3a_4 - a_3^3 - a_4^2 - a_2^2a_5 + a_3a_5. \tag{1.5}$$

Over the past, several authors established sharp bound of second-order Hankel determinants, see [1, 8]. However, it is more complicated to compute the bound of the third-order Hankel determinant, which is obtained by taking $q = 3$ in (1.4), and there are few articles concerning sharp bounds. Zaprawa [21] estimated that

$$|H_{3,1}(f)| \leq \begin{cases} 1, & f \in \mathcal{S}^* \\ 49/540, & f \in \mathcal{C}. \end{cases}$$

Later, for the class $\mathcal{S}^*$, Kwon et al. [11] obtained $|H_{3,1}(f)| \leq 8/9$, which is recently best improved to $4/9$ by Kowalczyk et al. [7]. Lecko et al. [12] has obtained $|H_{3,1}(f)| \leq 1/9$, which is sharp for functions belonging to the class $\mathcal{S}^*(1/2)$. For more on Hankel determinants, we refer [6, 7, 12, 18].

By specializing the function φ in $\mathcal{S}^*(\varphi)$ and $\mathcal{C}(\varphi)$, we obtain some well-known subclasses of $\mathcal{S}^*$ and $\mathcal{C}$, which are illustrated below:

(1) Janowski [4] introduced, $\mathcal{S}^*((1 + Az)/(1 + Bz)) =: \mathcal{S}^*[A, B]$ and $\mathcal{C}((1 + Az)/(1 + Bz)) =: \mathcal{C}[A, B]$ for $-1 \leq B < A \leq 1$.
(2) On taking $\varphi(z) = \sqrt{1 + z}$, Sokół and Stankiewicz [20] introduced the class, $\mathcal{S}_L^* := \mathcal{S}(\sqrt{1 + z})$ and $\mathcal{C}_L := \mathcal{C}(\sqrt{1 + z})$.

Similarly, the classes $\mathcal{S}_e^*$ and $\mathcal{C}_e$ were introduced and studied by Mendiratta et al. [15] by considering $\varphi(z) = e^z$ in (1.2) and (1.3) respectively, defined as:

$$\mathcal{S}_e^* = \left\{ f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec e^z \right\} \quad \text{and} \quad \mathcal{C}_e = \left\{ f \in \mathcal{A} : 1 + \frac{zf''(z)}{f'(z)} \prec e^z \right\}.$$

Several radius problems [15] and implications of differential subordination [16] have been established for various subclasses associated with the exponential function. Zaprawa [22] has improved the bounds of the third Hankel determinants related to the classes $\mathcal{S}_e^*$ and $\mathcal{C}_e$, which is given by

$$|H_{3,1}(f)| \leq \begin{cases} 0.385, & f \in \mathcal{S}_e^* \\ 0.021, & f \in \mathcal{C}_e. \end{cases} \tag{1.6}$$

However, the results were not sharp.

In the present investigation, we obtain the sharp bound of the result given in (1.6). We estimate sharp third hankel determinant, bounds of the sixth and seventh coefficients and fourth Hankel determinant for functions in $\mathcal{S}_e^*$ and $\mathcal{C}_e$, respectively in the subsequent sections.

2. Hankel Determinants for $\mathcal{S}_e^*$

2.1. **Preliminaries.** Let $f \in \mathcal{S}_e^*$, then there exists a Schwarz function $w(z)$ such that

$$\frac{zf'(z)}{f(z)} = e^{w(z)}. \tag{2.1}$$

Suppose that $p(z) = 1 + p_1z + p_2z^2 + \cdots \in \mathcal{P}$ and consider $w(z) = (p(z) - 1)/(p(z) + 1)$. Further, by substituting the expansions of $w(z)$, $p(z)$ and $f(z)$ in equation (2.1) and then comparing the coefficients, we obtain the expressions of $a_i (i = 2, 3, ..., 7)$ in terms of $p_j (j = 1, 2, ..., 5)$, given as follows:

$$a_2 = \frac{1}{2}p_1, \quad a_3 = \frac{1}{16}\left(4p_2 + p_1^2\right), \quad a_4 = \frac{1}{288}\left(- p_1^3 + 12p_1p_2 + 48p_3\right), \tag{2.2}$$

$$a_5 = \frac{1}{1152}\Big(p_1^4 - 12p_1^2p_2 + 24p_1p_3 + 144p_4\Big), \tag{2.3}$$

$$a_6 = \frac{1}{57600}\Big(- 17p_1^5 + 220p_1^3p_2 - 480p_1p_2^2 - 480p_1^2p_3 - 480p_2p_3 + 720p_1p_4 + 5760p_5\Big), \tag{2.4}$$

and

$$a_7 = \frac{1}{8294400}\Big(881p_1^6 - 13260p_1^4p_2 + 48240p_1^2p_2^2 - 14400p_2^3 + 29040p_1^3p_3 - 106560p_1p_2p_3 - 57600p_3^2 - 56160p_1^2p_4 - 86400p_2p_4 + 69120p_1p_5\Big). \tag{2.5}$$

The formula for p_i $(i = 2, 3, 4)$, which is included in the Lemma 2.1 below, plays a vital role in establishing the sharp bound for Hankel determinants and forms the foundation for our main results. See the recent survey article [2] for more information on the coefficients of the class $\mathcal{P}$.

Lemma 2.1. *[10, 13] Let $p \in \mathcal{P}$ has the form $1 + \sum_{n=1}^{\infty} p_n z^n$. Then*

$$p_2 = \frac{p_1^2 + \gamma(4 - p_1^2)}{2}, \tag{2.6}$$

$$p_3 = \frac{p_1^3 + 2p_1(4 - p_1^2)\gamma - p_1(4 - p_1^2)\gamma^2 + 2(4 - p_1^2)(1 - |\gamma|^2)\eta}{4}, \tag{2.7}$$

and

$$p_4 = \frac{p_1^4 + (4 - p_1^2)\gamma(p_1^2(\gamma^2 - 3\gamma + 3) + 4\gamma) - 4(4 - p_1^2)(1 - |\gamma|^2)(p_1(\gamma - 1)\eta + \bar{\gamma}\eta^2 - (1 - |\eta|^2)\rho)}{8}, \tag{2.8}$$

for some γ, η and ρ such that $|\gamma| \leq 1$, $|\eta| \leq 1$ and $|\rho| \leq 1$.

2.2. **Sharp Third Hankel Determinant for $\mathcal{S}_e^*$.** The following theorem presents the sharp bound for $|H_{3,1}(f)|$ for functions belonging to the class $\mathcal{S}_e^*$.

Theorem 2.2. *Let $f \in \mathcal{S}_e^*$. Then*

$$|H_{3,1}(f)| \leq 1/9. \tag{2.9}$$

This result is sharp.

Proof. Since the class $\mathcal{P}$ is invariant under rotation, the value of p_1 belongs to the interval [0,2]. Let $p := p_1$ and then substitute the values of $a_i (i = 2, 3, 4, 5)$ in equation (1.5) from equations (2.2) and (2.3). We get

$$H_{3,1}(f) = \frac{1}{331776}\Big(- 211p^6 + 420p^4p_2 - 1872p^2p_2^2 - 5184p_2^3 + 2544p^3p_3 + 10944pp_2p_3 - 9216p_3^2 - 7776p^2p_4 + 10368p_2p_4\Big).$$

After simplifying the calculations through (2.6)-(2.8), we obtain

$$H_{3,1}(f) = \frac{1}{331776}\Big(\beta_1(p, \gamma) + \beta_2(p, \gamma)\eta + \beta_3(p, \gamma)\eta^2 + \phi(p, \gamma, \eta)\rho\Big),$$

for $\gamma, \eta, \rho \in \mathbb{D}$. Here

$$\begin{aligned}
\beta_1(p,\gamma) :&= -13p^6 - 36\gamma^2p^2(4-p^2)^2 - 360\gamma^3p^2(4-p^2)^2 + 72\gamma^4p^2(4-p^2)^2 + 78\gamma p^4(4-p^2)\\
&\quad + 120p^4\gamma^2(4-p^2) - 324p^4\gamma^3(4-p^2) - 1296\gamma^2p^2(4-p^2),\\
\beta_2(p,\gamma) :&= 24(1-|\gamma|^2)(4-p^2)(17p^3 + 54\gamma p^3 + 30p\gamma(4-p^2) - 12p\gamma^2(4-p^2)),\\
\beta_3(p,\gamma) :&= 144(1-|\gamma|^2)(4-p^2)(-16(4-p^2) - 2|\gamma|^2(4-p^2) + 9p^2\bar{\gamma}),\\
\phi(p,\gamma,\eta) :&= 1296(1-|\gamma|^2)(4-p^2)(1-|\eta|^2)(2(4-p^2)\gamma - p^2).
\end{aligned}$$

By choosing $x = |\gamma|$, $y = |\eta|$ and utilizing the fact that $|\rho| \leq 1$, the above expression reduces to the following:

$$|H_{3,1}(f)| \leq \frac{1}{331776}\bigg(|\beta_1(p,\gamma)| + |\beta_2(p,\gamma)|y + |\beta_3(p,\gamma)|y^2 + |\phi(p,\gamma,\eta)|\bigg) \leq M(p,x,y),$$

where

$$M(p,x,y) = \frac{1}{331776}\bigg(m_1(p,x) + m_2(p,x)y + m_3(p,x)y^2 + m_4(p,x)(1-y^2)\bigg), \tag{2.10}$$

with

$$\begin{aligned}
m_1(p,x) :&= 13p^6 + 36x^2p^2(4-p^2)^2 + 360x^3p^2(4-p^2)^2 + 72x^4p^2(4-p^2)^2 + 78xp^4(4-p^2)\\
&\quad + 120p^4x^2(4-p^2) + 324p^4x^3(4-p^2) + 1296x^2p^2(4-p^2),\\
m_2(p,x) :&= 24(1-x^2)(4-p^2)(17p^3 + 54xp^3 + 30px(4-p^2) + 12px^2(4-p^2)),\\
m_3(p,x) :&= 144(1-x^2)(4-p^2)(16(4-p^2) + 2x^2(4-p^2) + 9p^2x),\\
m_4(p,x) :&= 1296(1-x^2)(4-p^2)(2x(4-p^2) + p^2).
\end{aligned}$$

In the closed cuboid $U : [0,2]\times[0,1]\times[0,1]$, we now maximise $M(p,x,y)$, by locating the maximum values in the interior of the six faces, on the twelve edges, and in the interior of U.

(1) We start by taking into account every internal point of U. Assume that $(p,x,y) \in (0,2)\times(0,1)\times(0,1)$. We calculate $\partial M/\partial y$ to identify the points of maxima in the interior of U. We get

$$\begin{aligned}
\frac{\partial M}{\partial y} = \frac{(4-p^2)(1-x^2)}{13824}\bigg(&24px(5+2x) + p^3(17+24x-12x^2) + 96(8-9x+x^2)y\\
&- 12p^2(25-27x+2x^2)y\bigg).
\end{aligned}$$

Now $\dfrac{\partial M}{\partial y} = 0$ gives

$$y = y_0 := \frac{p(17p^2 + 120x + 24p^2x + 48x^2 - 12p^2x^2)}{12(-64 + 25p^2 + 72x - 27p^2x - 8x^2 + 2p^2x^2)}.$$

The existence of critical points requires that y_0 belong to $(0,1)$, which is only possible when

$$\begin{aligned}
300p^2 + 864x + 24p^2x^2 > 17p^3 + 120px + 24p^3x + 48px^2 - 12p^3x^2\\
+ 768 + 864x + 24p^2x^2.
\end{aligned} \tag{2.11}$$

Now, we find the solution satisfying the inequality (2.11) for the existence of critical points using the hit and trial method. If we assume p tends to 0, then there does not exist any $x \in (0,1)$ satisfying the equation (2.11). But, when p tends to 2, the equation (2.11) holds for all $x < 37/54$. We also observe that there does not exist any $p \in (0,2)$ when $x \in (37/54, 1)$. Similarly, if we assume x tends to 0, then for all $p > 1.68218$, the equation (2.11) holds.

After calculations, we observe that there does not exist any $x \in (0,1)$ when $p \in (0, 1.68218)$. Thus, the domain for the solution of the equation is $(1.68218, 2) \times (0, 37/54)$. Now, we examine that $\frac{\partial M}{\partial y}|_{y=y_0} \neq 0$ in $(1.68218, 2) \times (0, 37/54)$. So, we conclude that the function M has no critical point in $(0,2) \times (0,1) \times (0,1)$.

(2) The interior of each of the cuboid U's six faces is now being considered.
On $p = 0, M(p, x, y)$ turns into

$$s_1(x,y) := \frac{(1-x^2)(8y^2 + x^2y^2 + 9x(1-y^2))}{72}, \quad x, y \in (0,1). \tag{2.12}$$

Since

$$\frac{\partial s_1}{\partial y} = \frac{(1-x^2)(x-1)(x-8)y}{36} \neq 0, \quad x, y \in (0,1),$$

indicates that s_1 has no critical points in $(0,1) \times (0,1)$.
On $p = 2, M(p, x, y)$ reduces to

$$M(2,x,y) := \frac{13}{5184}, \quad x, y \in (0,1). \tag{2.13}$$

On $x = 0, M(p, x, y)$ becomes

$$s_2(p,y) : = \frac{13p^6 + (4-p^2)(408p^3y + 2304y^2(4-p^2) + 1296p^2(1-y^2)}{331776} \tag{2.14}$$

with $p \in (0,2)$ and $y \in (0,1)$. To determine the points of maxima, we solve $\partial s_2/\partial p = 0$ and $\partial s_2/\partial y = 0$. After solving $\partial s_2/\partial y = 0$, we get

$$y = \frac{17p^3}{12(25p^2 - 64)} (=: y_p). \tag{2.15}$$

In order to have $y_p \in (0,1)$ for the given range of y, $p_0 := p > \approx 1.68218$ is required. Based on calculations, $\partial s_2/\partial p = 0$ gives

$$1728p - 864p^3 + 13p^5 + 816p^2y - 340p^4y - 7872py^2 + 2400p^3y^2 = 0. \tag{2.16}$$

After substituting equation (2.15) into equation (2.16), we have

$$21233664p - 27205632p^3 + 11472192p^5 - 1613016p^7 + 2700p^9 = 0. \tag{2.17}$$

A numerical calculation suggests that $p \approx 1.35596 \in (0,2)$ is the solution of (2.17). So, we conclude that s_2 does not have any critical point in $(0,2) \times (0,1)$.

On $x = 1, M(p, x, y)$ reforms into

$$s_3(p,y) := M(p,1,y) = \frac{12672p^2 - 2952p^4 - 41p^6}{331776}, \quad p \in (0,2). \tag{2.18}$$

While computing $\partial s_3/\partial p = 0$, $p_0 := p \approx 1.43461$ comes out to be the critical point. Undergoing simple calculations, s_3 achieves its maximum value ≈ 0.0398426 at p_0.

On $y = 0, M(p, x, y)$ can be viewed as

$$\begin{aligned} s_4(p,x) : = \frac{1}{331776}\Big(& 41472x(1-x^2) + 576p^2(9 - 36x + x^2 + 46x^3 + 2x^4) \\ & - 24p^4(54 - 121x - 8x^2 + 174x^3 + 24x^4) \\ & + p^6(13 - 78x - 84x^2 + 36x^3 + 72x^4)\Big). \end{aligned}$$

After undergoing further calculations such as,

$$\frac{\partial s_4}{\partial x} = \frac{1}{331776}\Big(-82944x^2 + 41472(1-x^2) + 576p^2(-36+2x+138x^2+8x^3) - 24p^4(-121-16x+522x^2+96x^3) + p^6(-78-168x + 108x^2+288x^3)\Big)$$

and

$$\frac{\partial s_4}{\partial p} = \frac{1}{331776}\Big(6p^5(13-78x-84x^2+36x^3+72x^4) - 96p^3(54-121x-8x^2 + 174x^3+24x^4) + 1152p(9-36x+x^2+46x^3+2x^4)\Big),$$

we observe that no solution in $(0,2)\times(0,1)$ exists of the system of equations $\partial s_4/\partial x = 0$ and $\partial s_4/\partial p = 0$.
On $y = 1, M(p,x,y)$ reduces to

$$\begin{aligned} s_5(p,x) := \frac{1}{331776}\Big(&2304px(5+2x-5x^2-2x^3) - 4608(-8+7x^2+x^4) \\ &+ 576p^2(-32+9x+38x^2+x^3+6x^4) - 24p^5(17+24x \\ &- 29x^2-24x^3+12x^4) + 96p^3(17-6x-41x^2+6x^3 \\ &+ 24x^4) - 24p^4(-96+41x+130x^2+12x^3+36x^4) \\ &+ p^6(13-78x-84x^2+36x^3+72x^4)\Big). \end{aligned}$$

The system of equations $\partial s_5/\partial x = 0$ and $\partial s_5/\partial p = 0$ also do not have any solution in $(0,2)\times(0,1)$.

(3) We next examine the maxima attained by $M(p,x,y)$ on the edges of the cuboid U. From equation (2.14), we have $M(p,0,0) = r_1(p) := (5184p^2 - 1296p^4 + 13p^6)/331776$. It is easy to observe that $r_1'(p) = 0$ whenever $p = \delta_0 := 0$ and $p = \delta_1 := 1.4367 \in [0,2]$ as its points of minima and maxima respectively. Hence,

$$M(p,0,0) \le 0.0159535, \quad p \in [0,2].$$

Now considering the equation (2.14) at $y = 1$, we get $M(p,0,1) = r_2(p) := (36864 - 18432p^2 + 1632p^3 + 2304p^4 - 408p^5 + 13p^6)/331776$. It is easy to observe that $r_2'(p) < 0$ in $[0,2]$ and hence $p = 0$ serves as the point of maxima. So,

$$M(p,0,1) \le \frac{1}{9}, \quad p \in [0,2].$$

Through computations, equation (2.14) shows that $M(0,0,y)$ attains its maxima at $y = 1$. This implies that

$$M(0,0,y) \le \frac{1}{9}, \quad y \in [0,1].$$

Since, the equation (2.18) does not involve x, we have $M(p,1,1) = M(p,1,0) = r_3(p) := (12672p^2 - 2952p^4 - 41p^6)/331776$. Now, $r_3'(p) = 4224p - 1968p^3 - 41p^5 = 0$ when $p = \delta_2 := 0$ and $p = \delta_3 := 1.43461$ in the interval $[0,2]$ with δ_2 and δ_3 as points of minima and maxima respectively. Hence

$$M(p,1,1) = M(p,1,0) \le 0.0398426, \quad p \in [0,2].$$

After considering $p = 0$ in (2.18), we get, $M(0, 1, y) = 0$. The equation (2.13) has no variables. So, on the edges, the maximum value of $M(p, x, y)$ is

$$M(2, 1, y) = M(2, 0, y) = M(2, x, 0) = M(2, x, 1) = \frac{13}{5184}, \quad x, y \in [0, 1].$$

Using equation (2.12), we obtain $M(0, x, 1) = r_4(x) := (8-7x^2-x^4)/72$. Upon calculations, we see that $r_4(x)$ is a decreasing function in $[0, 1]$ and attains its maxima at $x = 0$. Hence

$$M(0, x, 1) \leq \frac{1}{9}, \quad x \in [0, 1].$$

Again utilizing the equation (2.12), we get $M(0, x, 0) = r_5(x) := x(1 - x^2)/8$. On further calculations, we get $r_5'(x) = 0$ for $x = \delta_4 := 1/\sqrt{3}$. Also, $r_5(x)$ is an increases in $[0, \delta_4)$ and decreases in $(\delta_4, 1]$. So, it reaches its maximum value at δ_4. Thus

$$M(0, x, 0) \leq 0.0481125, \quad x \in [0, 1].$$

Given all the cases, the inequality (2.9) holds.
Let the function $f_1(z) \in \mathcal{S}_e^*$, be defined as

$$f_1(z) = z \exp\left(\int_0^z \frac{e^{t^3} - 1}{t} dt\right) = z + \frac{z^4}{3} + \frac{5z^7}{36} + \cdots,$$

with $f(0) = 0$ and $f'(0) = 1$, acts as an extremal function for the bound of $|H_{3,1}(f)|$ for $a_2 = a_3 = a_5 = 0$ and $a_4 = 1/3$.

∎

2.3. **Fourth Hankel Determinant for $\mathcal{S}_e^*$.** To continue, we state the next lemma, which will be utilized frequently.

Lemma 2.3. *[9, 19] Let $p = 1 + \sum_{n=1}^{\infty} p_n z^n \in \mathcal{P}$. Then*

$$|p_n| \leq 2, \quad n \geq 1,$$

$$|p_{n+k} - \nu p_n p_k| \leq \begin{cases} 2, & 0 \leq \nu \leq 1; \\ 2|2\nu - 1|, & otherwise, \end{cases}$$

and

$$|p_1^3 - \nu p_3| \leq \begin{cases} 2|\nu - 4|, & \nu \leq 4/3; \\ 2\nu\sqrt{\dfrac{\nu}{\nu - 1}}, & 4/3 < \nu. \end{cases}$$

We derive the expression of the fourth Hankel determinant when $q = 4$ and $n = 1$ are put into equation (1.4) as follows :

$$H_{4,1}(f) = a_7 H_{3,1}(f) - a_6 T_1 + a_5 T_2 - a_4 T_3, \tag{2.19}$$

where

$$T_1 := a_6(a_3 - a_2^2) + a_3(a_2 a_5 - a_3 a_4) - a_4(a_5 - a_2 a_4), \tag{2.20}$$

$$T_2 := a_3(a_3 a_5 - a_4^2) - a_5(a_5 - a_2 a_4) + a_6(a_4 - a_2 a_3), \tag{2.21}$$

and

$$T_3 := a_4(a_3 a_5 - a_4^2) - a_5(a_2 a_5 - a_3 a_4) + a_6(a_4 - a_2 a_3). \tag{2.22}$$

Now, using Lemma 2.3, we first determine the bounds of T_1, T_2, and T_3.
By substituting the values of a_i's $(i = 2, 3, ..., 6)$ in (2.20) using (2.2)-(2.4), we obtain

$$\begin{aligned}5529600T_1 &= 581p_1^7 + 5040p_1^4p_3 + 25920p_1^2p_2p_3 - 7068p_1^5p_2 + 11040p_1^3p_4 - 115200p_3p_4\\ &\quad + 7920p_1^3p_2^2 - 69120p_2^2p_3 + 74880p_1p_2p_4 - 25920p_1p_2^3 + 57600p_1p_3^2\\ &\quad + 138240p_2p_5 - 103680p_1^2p_5\end{aligned}$$

or

$$\begin{aligned}5529600|T_1| &\leq |p_1^4(581p_1^3 + 5040p_3)| + |p_1^2p_2(25920p_3 - 7068p_1^3)| + |p_4(11040p_1^3 - 115200p_3)|\\ &\quad + |p_2^2(7920p_1^3 - 69120p_3)| + |p_1p_2(74880p_4 - 25920p_2^2)| + |57600p_1p_3^2|\\ &\quad + |p_5(138240p_2 - 103680p_1^2)|. \qquad (2.23)\end{aligned}$$

Using Lemma 2.3 and the triangle inequality, we arrive at

$$|p_1^4(581p_1^3 + 5040p_3)| \leq 235648, \quad |p_1^2p_2(25920p_3 - 7068p_1^3)| \leq 4976640\sqrt{\frac{15}{1571}}, \qquad (2.24)$$

$$|p_4(11040p_1^3 - 115200p_3)| \leq 1843200\sqrt{\frac{15}{217}}, \quad |p_2^2(7920p_1^3 - 69120p_3)| \leq 442368\sqrt{\frac{30}{17}}, \qquad (2.25)$$

$$|p_1p_2(74880p_4 - 25920p_2^2)| \leq 599040, \quad |57600p_1p_3^2| \leq 460800, \qquad (2.26)$$

and

$$|p_5(138240p_2 - 103680p_1^2)| \leq 552960. \qquad (2.27)$$

By considering equation (2.23) in view of (2.24)-(2.27),

$$\begin{aligned}|T_1| &\leq \frac{1848448 + 4976640\sqrt{\frac{15}{1571}} + 1843200\sqrt{\frac{15}{217}} + 442368\sqrt{\frac{30}{17}}}{5529600}\\ &\approx 0.616137.\end{aligned}$$

Now, we calculate the bound of T_2 in the similar way by substituting the values of a_i's $(i = 2, 3, ..., 6)$ in (2.21) from equations (2.2)-(2.4), as follows:

$$\begin{aligned}22118400T_2 &= 235p_1^8 + 8712p_1^5p_3 + 37440p_1^3p_2p_3 - 1156p_1^6p_2 - 63360p_1p_2^2p_3 - 14640p_1^4p_2^2\\ &\quad + 161280p_1p_3p_4 - 8400p_1^4p_4 + 368640p_3p_5 - 76800p_1^3p_5 - 8640p_1^2p_2^3\\ &\quad + 172800p_2^2p_4 - 345600p_4^2 - 40320p_1^2p_3^2 - 184320p_2p_3^2 + 178560p_1^2p_2p_4\\ &\quad - 184320p_1p_2p_5\end{aligned}$$

or

$$\begin{aligned}22118400|T_2| &\leq |p_1^5(235p_1^3 + 8712p_3)| + |p_1^3p_2(37440p_3 - 1156p_1^3)| + | - p_1p_2^2(63360p_3 + 14640p_1^3)|\\ &\quad + |p_1p_4(161280p_3 - 8400p_1^3)| + |p_5(368640p_3 - 76800p_1^3)| + | - 8640p_1^2p_2^3|\\ &\quad + |p_4(172800p_2^2 - 345600p_4)| + | - p_3^2(184320p_2 + 40320p_1^2)|\\ &\quad + |p_1p_2(178560p_1p_4 - 184320p_5)|. \qquad (2.28)\end{aligned}$$

Lemma 2.3 and the triangle inequality lead us to

$$|p_1^5(235p_1^3 + 8712p_3)| \leq 617728, \quad |p_1^3p_2(37440p_3 - 1156p_1^3)| \leq 14376960\sqrt{\frac{65}{9071}}, \qquad (2.29)$$

$$|p_1p_2^2(63360p_3 + 14640p_1^3)| \leq 1950720, \quad |p_1p_4(161280p_3 - 8400p_1^3)| \leq 737280\sqrt{\frac{42}{13}}, \qquad (2.30)$$

$$|p_5(368640p_3 - 76800p_1^3)| \leq 2949120\sqrt{\frac{6}{19}}, \quad |8640p_1^2p_2^3| \leq 276480, \tag{2.31}$$

$$|p_4(172800p_2^2 - 345600p_4)| \leq 1382400, \quad |p_3^2(184320p_2 + 40320p_1^2)| \leq 2119680, \tag{2.32}$$

and

$$|p_1p_2(178560p_1p_4 - 184320p_5)| \leq 1474560. \tag{2.33}$$

By substituting the values from equations (2.29)-(2.33) in equation (2.28),

$$\begin{aligned}|T_2| &\leq \frac{7821568 + 14376960\sqrt{\frac{65}{9071}} + 2949120\sqrt{\frac{6}{19}} + 737280\sqrt{\frac{42}{13}}}{22118400}\\ &\approx 0.543487.\end{aligned}$$

Next, we determine the bound of T_3, by replacing the values of a_i's $(i = 2, 3, ..., 6)$ from equations (2.2)-(2.4) in (2.22), as follows:

$$\begin{aligned}597196800T_3 = {} & 6120p_1^8 + 143424p_1^5p_3 - 425p_1^9 - 9000p_1^6p_3 + 9000p_1^7p_2 + 172800p_1^4p_2p_3\\ & + 302400p_1^3p_2^2 - 2764800p_3^3 + 1036800p_1^3p_2p_4 + 6220800p_2p_3p_4 - 17280p_1^4p_2^2\\ & + 9953280p_3p_5 - 2073600p_1^3p_5 + 967680p_1^3p_2p_3 - 64512p_1^6p_2 - 1036800p_1p_2p_3^2\\ & - 32400p1^5p_2^2 - 777600p_1^2p_2^2p_3 + 1244160p_1p_3p_4 - 259200p_1^4p_4 - 97200p_1^5p_4\\ & + 1555200p_1p_2^2p_4 - 4665600p_1p_4^2 - 414720p_1p_2^2p_3 - 4976640p_1p_2p_5 - 829440p_2p_3^2\\ & - 829440p_1^2p_3^2 + 414720p_1^2p_2^3 - 622080p_1^2p_2p_4 - 172800p_1^3p_2^3\end{aligned}$$

or

$$\begin{aligned}597196800|T_3| \leq {} & |p_1^5(6120p_1^3 + 143424p_3)| + |p_1^6(425p_1^3 + 9000p_3)| + |p_1^4p_2(9000p_1^3 + 172800p_3)|\\ & + |p_3^2(302400p_1^3 - 2764800p_3)| + |p_2p_4(1036800p_1^3 + 6220800p_3)| + |17280p_1^4p_2^2|\\ & + |p_5(9953280p_3 - 2073600p_1^3)| + |p_1^3p_2(967680p_3 - 64512p_1^3)| + |1036800p_1p_2p_3^2|\\ & + |p_1^2p_2^2(32400p1^3 + 777600p_3)| + |p_1p_4(1244160p_3 - 259200p_1^3)| + |97200p_1^5p_4|\\ & + |p_1p_4(1555200p_2^2 - 4665600p_4)| + |p_1p_2(414720p_2p_3 + 4976640p_5)|\\ & + |p_3^2(829440p_2 + 829440p_1^2)| + |p_1^2p_2(414720p_2^2 - 622080p_4)| + |172800p_1^3p_2^3|.\end{aligned} \tag{2.34}$$

By applying Lemma 2.3 and the triangle inequality,

$$|p_1^5(6120p_1^3 + 143424p_3)| \leq 10745856, \quad |p_1^6(425p_1^3 + 9000p_3)| \leq 1369600, \tag{2.35}$$

$$|p_1^4p_2(9000p_1^3 + 172800p_3)| \leq 13363200, \quad |p_3^2(302400p_1^3 - 2764800p_3)| \leq 58982400\sqrt{\frac{3}{19}}, \tag{2.36}$$

$$|p_2p_4(1036800p_1^3 + 6220800p_3)| \leq 82944000, \quad |p_5(9953280p_3 - 2073600p_1^3)| \leq 79626240\sqrt{\frac{6}{19}}, \tag{2.37}$$

$$|p_1^3p_2(967680p_3 - 64512p_1^3)| \leq 2211840\sqrt{210}, \quad |p_1^2p_2^2(32400p1^3 + 777600p_3)| \leq 29030400, \tag{2.38}$$

$$|p_1p_4(1244160p_3 - 259200p_1^3)| \leq 19906560\sqrt{\frac{6}{19}}, \quad |p_1p_4(1555200p_2^2 - 4665600p_4)| \leq 37324800, \tag{2.39}$$

$$|p_1p_2(414720p_2p_3+4976640p_5)|\leq 46448640,\quad |p_3^2(829440p_2+829440p_1^2)|\leq 19906560, \tag{2.40}$$

$$|17280p_1^4p_2^2|+|1036800p_1p_2p_3^2|+|97200p_1^5p_4|+|172800p_1^3p_2^3|\leq 34974720, \tag{2.41}$$

and

$$|p_1^2p_2(414720p_2^2-622080p_4)|\leq 9953280. \tag{2.42}$$

So, by using equations (2.35)-(2.42) in equation (2.34), we get

$$\begin{aligned}|T_3|&\leq \frac{286061056+58982400\sqrt{\frac{3}{19}}+99532800\sqrt{\frac{6}{19}}+2211840\sqrt{210}}{597196800}\\&\approx 0.665582.\end{aligned}$$

Remark 2.4. On the basis of the above calculations, the bounds of T_1, T_2 and T_3 are 0.616137, 0.543487 and 0.665582 respectively.

Now, to proceed ahead, we seek the bounds of initial coefficients a_i for $i=2,3,4,5$. These bounds have been derived in [22], presented below in the form of the following remark.

Remark 2.5. For $f\in\mathcal{S}_e^*$, $|a_2|\leq 1$, $|a_3|\leq 3/4$, $|a_4|\leq 17/36$ and $|a_5|\leq 25/72$. Here the first three bounds are sharp.

It is significantly more difficult to find coefficient boundaries for $n>5$. To solve this problem, we use the Lemma 2.3 to establish the bounds for the sixth and seventh coefficient for functions of the class $\mathcal{S}_e^*$ in the following Lemma.

Lemma 2.6. *Let $f\in\mathcal{S}_e^*$. Then $|a_6|\leq 587/1800\approx 0.326111$ and $|a_7|\leq 1397/4320\approx 0.32338$.*

Proof. By suitably rearranging the terms given in equation (2.4), we have

$$57600a_6=220p_1^3p_2-480p_1^2p_3-480p_1p_2^2+720p_1p_4-17p_1^5-480p_2p_3+5760p_5.$$

Using triangle inequality, it can be viewed as

$$\begin{aligned}57600|a_6|\leq{}&|p_1^2(220p_1p_2-480p_3)|+|p_1(720p_4-480p_2^2)|+|-17p_1^5|\\&+|5760p_5-480p_2p_3|.\end{aligned} \tag{2.43}$$

Using Lemma 2.3, we arrive at the following inequalities:

$$|p_1^2(220p_1p_2-480p_3)|\leq 3840,\quad |p_1(720p_4-480p_2^2)|\leq 2880,\quad |17p_1^5|\leq 544, \tag{2.44}$$

and

$$|5760p_5-480p_2p_3|\leq 11520. \tag{2.45}$$

By using equation (2.44) and (2.45) in equation (2.43), we obtain

$$|a_6|\leq\frac{587}{1800}\approx 0.326111.$$

Similarly, considering equation (2.5), we have

$$\begin{aligned}8294400a_7={}&881p_1^6-13260p_1^4p_2+48240p_1^2p_2^2-14400p_2^3+29040p_1^3p_3-56160p_1^2p_4\\&+69120p_1p_5-106560p_1p_2p_3-57600p_3^2-86400p_2p_4.\end{aligned}$$

Through the triangle inequality, it can also be seen as

$$\begin{aligned}8294400|a_7|\leq{}&|p_1^4(881p_1^2-13260p_2)|+|p_2^2(48240p_1^2-14400p_2)|+|p_1(69120p_5-106560p_2p_3)|\\&+|p_1^2(29040p_1p_3-56160p_4)|+|-57600p_3^2|+|-86400p_2p_4|.\end{aligned} \tag{2.46}$$

Lemma 2.3 takes us to the following:

$$|p_1^4(881p_1^2-13260p_2)|\leq 424320,\quad |p_2^2(48240p_1^2-14400p_2)|\leq 656640, \tag{2.47}$$

$$|p_1(69120p_5 - 106560p_2p_3)| \leq 576000, \quad |p_1^2(29040p_1p_3 - 56160p_4)| \leq 449280, \tag{2.48}$$

and

$$|57600p_3^2| + |86400p_2p_4| \leq 576000. \tag{2.49}$$

By substituting the values from equations (2.47)-(2.49) in (2.46), we have

$$|a_7| \leq \frac{1397}{4320} \approx 0.32338.$$

∎

Theorem 2.7. *Let $f \in \mathcal{S}_e^*$. Then*

$$|H_{4,1}(f)| \leq 0.29059.$$

The proof of the above theorem follows by substituting the values obtained from Theorem 2.2, Remark 2.4, Remark 2.5 and Lemma 2.6 in the equation (2.19), therefore, it is skipped here.

3. Hankel Determinants for $\mathcal{C}_e$

3.1. **Preliminaries.** When $f \in \mathcal{C}_e$, we replace the L.H.S of equation (2.1) by $1 + zf''(z)/f'(z)$ and arrive at the following equation

$$1 + \frac{zf''(z)}{f'(z)} = e^{w(z)}.$$

Proceeding on the similar lines as done for the class $\mathcal{S}_e^*$, we obtain $a_i (i = 2, 3, ..., 7)$ in terms of $p_j (j = 1, 2, ..., 5)$, then compare the corresponding coefficients as follows:

$$a_2 = \frac{1}{4}p_1, \quad a_3 = \frac{1}{48}\Big(p_1^2 + 4p_2\Big), \quad a_4 = \frac{1}{1152}\Big(-p_1^3 + 12p_1p_2 + 48p_3\Big), \tag{3.1}$$

$$a_5 = \frac{1}{5760}\Big(p_1^4 - 12p_1^2p_2 + 24p_1p_3 + 144p_4\Big), \tag{3.2}$$

$$a_6 = \frac{1}{345600}\Big(-17p_1^5 + 220p_1^3p_2 - 480p_1p_2^2 - 480p_1^2p_3 - 480p_2p_3 + 720p_1p_4 + 5760p_5\Big), \tag{3.3}$$

and

$$a_7 = \frac{1}{58060800}\Big(881p_1^6 - 13260p_1^4p_2 + 48240p_1^2p_2^2 - 14400p_2^3 + 29040p_1^3p_3 - 106560p_1p_2p_3 - 57600p_3^2 - 56160p_1^2p_4 - 86400p_2p_4 + 69120p_1p_5\Big). \tag{3.4}$$

3.2. **Sharp Third Hankel Determinant for $\mathcal{C}_e$.** In this section, we establish the best bound of $|H_{3,1}(f)|$ for functions that belong to the class $\mathcal{C}_e$.

Theorem 3.1. *Let $f \in \mathcal{C}_e$. Then*

$$|H_{3,1}(f)| \leq \frac{1}{144}. \tag{3.5}$$

This bound is sharp.

Proof. We follow the same steps which were used to prove Theorem 2.2. The values of $a_i's(i = 2, 3, 4, 5)$ from equations (3.1) and (3.2) are substituted into equation (1.5). Thus

$$H_{3,1}(f) = \frac{1}{6635520}\bigg(- 173p^6 + 552p^4p_2 - 1872p^2p_2^2 - 3840p_2^3 + 2208p^3p_3 + 8064pp_2p_3 - 11520p_3^2 - 6912p^2p_4 + 13824p_2p_4\bigg).$$

Using (2.6)-(2.8) for simplification, we arrive at

$$H_{3,1}(f) = \frac{1}{6635520}\bigg(\alpha_1(p,\gamma) + \alpha_2(p,\gamma)\eta + \alpha_3(p,\gamma)\eta^2 + \psi(p,\gamma,\eta)\rho\bigg),$$

where $\gamma, \eta, \rho \in \mathbb{D}$,

$$\begin{aligned}
\alpha_1(p,\gamma) :&= -5p^6 - 180\gamma^2p^2(4-p^2)^2 + 1536\gamma^3(4-p^2)^2 - 240\gamma^3p^2(4-p^2)^2 + 144\gamma^4p^2(4-p^2)^2 \\
&\quad + 12\gamma p^4(4-p^2) - 120p^4\gamma^2(4-p^2), \\
\alpha_2(p,\gamma) :&= (1-|\gamma|^2)(4-p^2)(240p^3 - 288p\gamma(4-p^2) - 576p\gamma^2(4-p^2)), \\
\alpha_3(p,\gamma) :&= (1-|\gamma|^2)(4-p^2)(-2880(4-p^2) - 576|\gamma|^2(4-p^2)), \\
\psi(p,\gamma,\eta) :&= 3456\gamma(1-|\gamma|^2)(4-p^2)^2(1-|\eta|^2).
\end{aligned}$$

Since $|\rho| \leq 1$, also for the simplicity of the calculations, assume $x = |\gamma|$ and $y = |\eta|$,

$$|H_{3,1}(f)| \leq \frac{1}{6635520}\bigg(|\alpha_1(p,\gamma)| + |\alpha_2(p,\gamma)|y + |\alpha_3(p,\gamma)|y^2 + |\psi(p,\gamma,\eta)|\bigg) \leq N(p,x,y),$$

where

$$N(p,x,y) = \frac{1}{6635520}\bigg(n_1(p,x) + n_2(p,x)y + n_3(p,x)y^2 + n_4(p,x)(1-y^2)\bigg), \tag{3.6}$$

with

$$\begin{aligned}
n_1(p,x) :&= 5p^6 + 180x^2p^2(4-p^2)^2 + 1536x^3(4-p^2)^2 + 240x^3p^2(4-p^2)^2 + 144x^4p^2(4-p^2)^2 \\
&\quad + 12xp^4(4-p^2) + 120p^4x^2(4-p^2), \\
n_2(p,x) :&= (1-x^2)(4-p^2)(240p^3 + 288px(4-p^2) + 576px^2(4-p^2)), \\
n_3(p,x) :&= (1-x^2)(4-p^2)(2880(4-p^2) + 576x^2(4-p^2)), \\
n_4(p,x) :&= 3456x(1-x^2)(4-p^2)^2.
\end{aligned}$$

We must maximise $N(p,x,y)$ in the closed cuboid $V : [0,2] \times [0,1] \times [0,1]$. By identifying the maximum values on the twelve edges, the interior of V, and the interiors of the six faces, we can prove this.

(1) We start by taking into account, every interior point of V. Assume that $(p,x,y) \in (0,2) \times (0,1) \times (0,1)$. We partially differentiate equation (3.6) with respect to y to locate the points of maxima in the interior of V. We obtain

$$\frac{\partial N}{\partial y} = \frac{(1-x^2)(4-p^2)}{138240}\bigg(24px(1+2x) - p^3(-5+6x+12x^2) + 96(5-6x+x^2)y - 24p^2(5-6x+x^2)y\bigg).$$

Now $\dfrac{\partial N}{\partial y} = 0$ gives

$$y = y_1 := \frac{5p^3 + 6px(4-p^2)(1+2x)}{24(4-p^2)(6x-x^2-5)}.$$

Since y_1 must be a member of $(0,1)$ for critical points to exist, this is only possible if

$$24(20+(p-24)x+(4+2p-p^2)x^2)+p^3(5-6x-12x^2)<24p^2(5-6x). \tag{3.7}$$

Now, we find the solutions satisfying the inequality (3.7) for the existence of critical points using the hit and trial method. If we assume p tends to 0 and 2, then no such $x\in(0,1)$ exists satisfying equation (3.7). Similarly, if we take x tending to 0 and 1, then there does not exist any $p\in(0,2)$ satisfying equation (3.7). Therefore, we conclude that the function N has no critical point in $(0,2)\times(0,1)\times(0,1)$.

(2) Now, we study the interior of each of the six faces of the cuboid V.
When $p=0, N(p,x,y)$ becomes

$$c_1(x,y):=\frac{y^2(15-12x^2-3x^4)+18x(1-y^2)-2x^3(5-9y^2)}{2160},\quad x,y\in(0,1). \tag{3.8}$$

Since

$$\frac{\partial c_1}{\partial y}=\frac{y(1-x)^2(x+1)(5-x)}{360}\neq 0,\quad x,y\in(0,1),$$

we note that, in $(0,1)\times(0,1)$, c_1 does not have any critical point.
When $p=2, N(p,x,y)$ settles into

$$N(2,x,y):=\frac{1}{20736},\quad x,y\in(0,1). \tag{3.9}$$

When $x=0, N(p,x,y)$ turns into

$$c_2(p,y):=\frac{(p^3+96y-24p^2y)^2}{1327104},\quad p\in(0,2)\quad\text{and}\quad y\in(0,1). \tag{3.10}$$

We solve $\partial c_2/\partial p=0$ and $\partial c_2/\partial y=0$ to locate the points of maxima. On solving $\partial c_2/\partial y=0$, we obtain

$$y=-\frac{p^3}{24(4-p^2)}(=:y_p).$$

Upon calculations, we observe that such y_p does not belong to $(0,1)$. Consequently, no such critical point of c_2 exists in $(0,2)\times(0,1)$.
When $x=1, N(p,x,y)$ becomes

$$N(p,1,y)=c_3(p,y):=\frac{24576-3264p^2-2448p^4+437p^6}{6635520},\quad p\in(0,2). \tag{3.11}$$

And while computing $\partial c_3/\partial p=0$, we notice that c_3 has no critical point in $(0,2)$.
When $y=0, N(p,x,y)$ reduces to

$$\begin{aligned}c_4(p,x):=\frac{1}{6635520}\Big(&6144x(9-5x^2)+192p^2x(-144+15x+100x^2+12x^3)\\&-48p^4x(-73+20x+80x^2+24x^3)\\&+p^6(5-12x+60x^2+240x^3+144x^4)\Big).\end{aligned}$$

Calculations lead to,

$$\begin{aligned}\frac{\partial c_4}{\partial x}=\frac{1}{6635520}\Big(&-61440x^2-6144(-9+5x^2)+192p^2x(15+200x+36x^2)\\&-48p^4x(20+160x+72x^2)+192p^2(-144+15x+100x^2\\&+12x^3)-48p^4(-73+20x+80x^2+24x^3)+p^6(-12+120x\\&+720x^2+576x^3)\Big)\end{aligned}$$

and

$$\frac{\partial c_4}{\partial p} = \frac{1}{6635520}\Big(384px(-144+15x+100x^2+12x^3) - 192p^3x(-73+20x + 80x^2+24x^3) + 6p^5(5-12x+60x^2+240x^3+144x^4)\Big).$$

No solution exist for the system of equations, $\partial c_4/\partial x = 0$ and $\partial c_4/\partial p = 0$, according to a numerical calculation, in $(0,2)\times(0,1)$.

When $y = 1$, $N(p,x,y)$ reduces to

$$\begin{aligned} c_5(p,x) := \frac{1}{6635520}\Big(&5p^6 + (4-p^2)(12p^4x + 120p^4x^2 + 180p^2(4-p^2)x^2 \\ &+ 1536(4-p^2)x^3 + 240p^2(4-p^2)x^3 + 144p^2(4-p^2)x^4 \\ &+ 3456(4-p^2)x(1-x^2) + 48(1-x^2)(p^3(5-6x \\ &- 12x^2) + 24px(1+2x)))\Big). \end{aligned}$$

The two equations $\partial c_5/\partial x = 0$ and $\partial c_5/\partial p = 0$ also do not assume any solution in $(0,2)\times(0,1)$.

(3) Next, we check the maximum values of $N(p,x,y)$ obtained on the edges of the cuboid V. From equation (3.10), we have $N(p,0,0) = t_1(p) := p^6/1327104$. It is easy to observe that $t_1'(p) = 0$ for $p = 0$ in the interval $[0,2]$. The maximum value of $t_1(p)$ is 0. Now the equation (3.10) reduces to $N(p,0,1) = t_2(p) := (96-24p^2+p^3)^2/1327104$ at $y = 1$. Since, $t_2'(p) < 0$ in $[0,2]$, hence $p = 0$ is the point of maxima. Thus

$$N(p,0,1) \le \frac{1}{144}, \quad p \in [0,2].$$

Through computations, equation (3.10) shows that $N(0,0,y)$ attains its maxima at $y = 1$. Hence

$$N(0,0,y) \le \frac{1}{144}, \quad y \in [0,1].$$

Since, the equation (3.11) is free from x, we have $N(p,1,1) = N(p,1,0) = t_3(p) := (24576 - 3264p^2 - 2448p^4 + 437p^6)/6635520$. Now, we observe that $t_3'(p) < 0$ in $[0,2]$, consequently, $t_3(p)$ attains its maximum at $p = 0$. Hence

$$N(p,1,1) = N(p,1,0) \le 0.0037037, \quad p \in [0,2].$$

On substituting $p = 0$ in equation (3.11), we get, $N(0,1,y) = 1/270$. The equation (3.9) does not contain any variable such as p, x and y. Therefore, the maxima of $N(p,x,y)$ on the edges is given by

$$N(2,1,y) = N(2,0,y) = N(2,x,0) = N(2,x,1) = \frac{1}{20736}, \quad x,y \in [0,1].$$

Using equation (3.8), we obtain $N(0,x,1) = t_4(x) := (15 - 12x^2 + 8x^3 - 3x^4)/2160$. Upon calculations, we see that t_4 is a decreasing function in $[0,1]$ and its maximum value is achieved at $x = 0$. Hence

$$N(0,x,1) \le \frac{1}{144}, \quad x \in [0,1].$$

On again using equation (3.8), we get $N(0,x,0) = t_5(x) := x(9-5x^2)/1080$. On further calculations, we get $t_5'(x) = 0$ for $x = \beta_0 := \sqrt{3/5}$. Also, $t_5(x)$ increases in $[0,\beta_0)$ and

decreases in $(\beta_0, 1]$. So, β_0 is the point of maxima. Thus

$$N(0, x, 0) \leq 0.00430331, \quad x \in [0, 1].$$

Because of all the cases discussed above, the inequality (3.5) holds.
The function $f_2(z) \in \mathcal{C}_e$, defined as

$$f_2(z) = \int_0^z \left(\exp\left(\int_0^y \frac{e^{t^3} - 1}{t} dt \right) \right) dy = z + \frac{z^4}{12} + \frac{5z^7}{252} + \cdots,$$

with $f(0) = f'(0) - 1 = 0$, plays the role of an extremal function for the bounds of $|H_{3,1}(f)|$ having values $a_3 = a_5 = 0$ and $a_4 = 1/12$. ∎

3.3. **Fourth Hankel Determinant for $\mathcal{C}_e$.** By selecting $q = 4$ and $n = 1$ in the equation (1.4), the expression of $|H_{4,1}(f)|$ can be obtained for functions in the class $\mathcal{C}_e$, which is given as follows:

$$H_{4,1}(f) = a_7 H_{3,1}(f) - a_6 U_1 + a_5 U_2 - a_4 U_3. \tag{3.12}$$

Here

$$U_1 := a_6(a_3 - a_2^2) + a_3(a_2 a_5 - a_3 a_4) - a_4(a_5 - a_2 a_4), \tag{3.13}$$

$$U_2 := a_3(a_3 a_5 - a_4^2) - a_5(a_5 - a_2 a_4) + a_6(a_4 - a_2 a_3), \tag{3.14}$$

and

$$U_3 := a_4(a_3 a_5 - a_4^2) - a_5(a_2 a_5 - a_3 a_4) + a_6(a_4 - a_2 a_3). \tag{3.15}$$

We start by determining the bounds for U_1, U_2, and U_3.
By substituting the values of a_i's $(i = 2, 3, ..., 6)$ in (3.13) from equations (3.1)-(3.3), we obtain

$$\begin{aligned} 132710400 U_1 = {} & 487 p_1^7 - 6304 p_1^5 p_2 + 11440 p_1^3 p_2^2 - 24960 p_1 p_2^3 + 5280 p_1^4 p_3 + 34560 p_1 p_3^2 \\ & + 19200 p_1^2 p_2 p_3 - 53760 p_2^2 p_3 + 57600 p_1 p_2 p_4 - 138240 p_3 p_4 \\ & + 184320 p_2 p_5 - 92160 p_1^2 p_5 + 8640 p_1^3 p_4, \end{aligned}$$

can also be viewed as the following due to the triangle inequality,

$$\begin{aligned} 132710400 |U_1| \leq {} & |p_1^5(487 p_1^2 - 6304 p_2)| + |p_1 p_2^2(11440 p_1^2 - 24960 p_2)| + |p_1 p_3(5280 p_1^3 + 34560 p_3)| \\ & + |p_2 p_3(19200 p_1^2 - 53760 p_2)| + |p_4(57600 p_1 p_2 - 138240 p_3)| \\ & + |p_5(184320 p_2 - 92160 p_1^2)| + |8640 p_1^3 p_4|. \end{aligned} \tag{3.16}$$

Using Lemma 2.3, we arrive at

$$|p_1^5(487 p_1^2 - 6304 p_2)| \leq 40320, \quad |p_1 p_2^2(11440 p_1^2 - 24960 p_2)| \leq 399360, \tag{3.17}$$

$$|p_1 p_3(5280 p_1^3 + 34560 p_3)| \leq 445440, \quad |p_2 p_3(19200 p_1^2 - 53760 p_2)| \leq 430080, \tag{3.18}$$

$$|p_4(57600 p_1 p_2 - 138240 p_3)| \leq 55296, \quad |p_5(184320 p_2 - 92160 p_1^2)| \leq 73728, \tag{3.19}$$

and

$$|8640 p_1^3 p_4| \leq 138240. \tag{3.20}$$

By substituting the values obtained from equations (3.17)-(3.20) in (3.16),

$$|U_1| \leq \frac{4121}{345600} \approx 0.0119242.$$

We replace the values of a_i's $(i = 2, 3, ..., 6)$ from equations (3.1)-(3.4) in equation (3.14) and proceed on the same lines to obtain the bound of U_2

$$\begin{aligned}1592524800U_2 = {} & 463p_1^8 - 2732p_1^6p_2 - 23472p_1^4p_2^2 - 14400p_1^2p_2^3 + 14592p_1^5p_3 - 108288p_1^2p_3^2 \\ & + 92928p_1^3p_2p_3 - 138240p_1p_2^2p_3 + 373248p_1^2p_2p_4 - 25344p_1^4p_4 + 276480p_2^2p_4 \\ & - 995328p_4^2 + 1105920p_3p_5 - 276480p_1p_2p_5 + 221184p_1p_3p_4 \\ & - 161280p_1^3p_5 - 322560p_2p_3^2,\end{aligned}$$

by implementing the triangle inequality,

$$\begin{aligned}1592524800|U_2| \leq {} & |p_1^6(463p_1^2 - 2732p_2)| + |p_1^2p_2^2(-23472p_1^2 - 14400p_2)| + |p_1^2p_3(14592p_1^3 - 108288p_3)| \\ & + |p_1p_2p_3(92928p_1^2 - 138240p_2)| + |p_1^2p_4(373248p_2 - 25344p_1^2)| \\ & + |p_4(276480p_2^2 - 995328p_4)| + |p_5(1105920p_3 - 276480p_1p_2)| \\ & + |221184p_1p_3p_4| + |-161280p_1^3p_5| + |-322560p_2p_3^2|. \qquad (3.21)\end{aligned}$$

By applying Lemma 2.3, we have

$$|p_1^6(463p_1^2 - 2732p_2)| \leq 349696, \quad |p_1^2p_2^2(23472p_1^2 + 14400p_2)| \leq 1963008, \qquad (3.22)$$

$$|p_1^2p_3(14592p_1^3 - 108288p_1^2p_3^2| \leq 866304\sqrt{\frac{282}{61}}, \quad |p_1p_2p_3(92928p_1^2 - 138240p_2)| \leq 2211840, \qquad (3.23)$$

$$|p_1^2p_4(373248p_2 - 25344p_1^2)| \leq 5971968, \quad |p_4(276480p_2^2 - 995328p_4)| \leq 3981312, \qquad (3.24)$$

$$|p_5(1105920p_3 - 276480p_1p_2)| \leq 4423680, \qquad (3.25)$$

and

$$|221184p_1p_3p_4| + |161280p_1^3p_5| + |322560p_2p_3^2| \leq 6045696. \qquad (3.26)$$

By substituting the values from equations (3.22)-(3.26) in (3.21),

$$|U_2| \leq \frac{24947200 + 866304\sqrt{\frac{282}{61}}}{1592524800} \approx 0.0168348.$$

Again, substitute the values of a_i's $(i = 2, 3, ..., 6)$ from equations (3.1)-(3.4) in (3.15) and proceed to calculate the bound of U_3 in the same manner.

$$\begin{aligned}38220595200U_3 = {} & 11424p_1^8 - 128256p_1^6p_2 + 10812p_1^7p_2 - 503p_1^9 + 69120p_1^4p_2^2 + 552960p_1^2p_2^3 \\ & - 42192p_1^5p_2^2 - 181440p_1^3p_2^3 + 206208p_1^4p_2p_3 - 11664p_1^6p_3 + 1889280p_1^3p_2p_3 \\ & - 1658880p_1p_2^2p_3 - 2211840p_1^2p_3^2 - 2211840p_2p_3^2 + 283392p_1^3p_3^2 - 967680p_1p_2p_3^2 \\ & + 3317760p_1p_3p_4 - 483840p_1^4p_4 + 1271808p_1^3p_2p_4 - 117504p_1^5p_4 \\ & + 1658880p_1p_2^2p_4 - 5971968p_1p_4^2 + 6635520p_2p_3p_4 - 331776p_1^2p_3p_4 \\ & + 26542080p_3p_5 - 6635520p_1p_2p_5 + 244224p_1^5p_3 - 794880p_1^2p_2^2p_3 \\ & - 2764800p_3^3 - 829440p_1^2p_2p_4 - 3870720p_1^3p_5,\end{aligned}$$

can be visualized as the following with the help of the triangle inequality,

$$\begin{aligned}38220595200|U_3| \leq\ & |p_1^6(11424p_1^2-128256p_2)|+|p_1^7(10812p_2-503p_1^2)|+|p_1^2p_2^2(69120p_1^2+552960p_2)|\\ &+|p_1^3p_2^2(-42192p_1^2-181440p_2)|+|p_1^4p_3(206208p_2-11664p_1^2)|\\ &+|p_1p_2p_3(1889280p_1^2-1658880p_2)|+|p_3^2(-2211840p_1^2-2211840p_2)|\\ &+|p_1p_3^2(283392p_1^2-967680p_2)|+|p_1p_4(3317760p_3-483840p_1^3)|\\ &+|p_1^3p_4(1271808p_2-117504p_1^2)|+|p_1p_4(1658880p_2^2-5971968p_4)|\\ &+|p_3p_4(6635520p_2-331776p_1^2)|+|p_5(26542080p_3-6635520p_1p_2)|\\ &+|244224p_1^5p_3-794880p_1^2p_2^2p_3-2764800p_3^3-829440p_1^2p_2p_4-3870720p_1^3p_5|.\end{aligned} \tag{3.27}$$

By applying Lemma 2.3, we get

$$|p_1^6(11424p_1^2-128256p_2)|\leq 16416768,\quad |p_1^7(10812p_2-503p_1^2)|\leq 2767872 \tag{3.28}$$

$$|p_1^2p_2^2(69120p_1^2+552960p_2)|\leq 22118400,\quad |p_1^3p_2^2(-42192p_1^2-181440p_2)|\leq 17012736, \tag{3.29}$$

$$|p_1^4p_3(206208p_2-11664p_1^2)|\leq 13197312,\quad |p_1p_2p_3(1889280p_1^2-1658880p_2)|\leq 33914880, \tag{3.30}$$

$$|p_3^2(2211840p_1^2+2211840p_2)|\leq 5308416,\quad |p_1p_3^2(283392p_1^2-967680p_2)|\leq 15482880, \tag{3.31}$$

$$|p_1p_4(3317760p_3-483840p_1^3)|\leq 106168320\sqrt{\frac{3}{41}},\quad |p_1^3p_4(1271808p_2-11750p_1^2)|\leq 40697856, \tag{3.32}$$

$$|p_1p_4(1658880p_2^2-5971968p_4)|\leq 47775744,\quad |p_3p_4(6635520p_2-331776p_1^2)|\leq 53084160, \tag{3.33}$$

$$|p_5(26542080p_3-6635520p_1p_2)|\leq 106168320, \tag{3.34}$$

and

$$|244224p_1^5p_3|+|794880p_1^2p_2^2p_3|+|2764800p_3^3|+|829440p_1^2p_2p_4|+|3870720p_1^3p_5|\leq 138387456. \tag{3.35}$$

So, by using equations (3.28)-(3.35) in (3.27), we get

$$|U_3|\leq \frac{560108544+106168320\sqrt{\frac{3}{41}}}{38220595200}\approx 0.015406.$$

Remark 3.2. The bounds of U_1, U_2 and U_3, based on the above calculations, are 0.0119242, 0.0168348, and 0.015406 respectively.

The bounds of a_i $(i=2,3,4,5)$ for functions in the class $\mathcal{C}_e$ are obtained in [22], presented below in the following remark:

Remark 3.3. For $f\in\mathcal{C}_e$, $|a_2|\leq 1/2$, $|a_3|\leq 1/4$, $|a_4|\leq 17/144$ and $|a_5|\leq 5/72$. The first three bounds are sharp.

Next, we calculate the bounds of the sixth and seventh coefficient of functions belonging to the class $\mathcal{C}_e$ to establish our main result along the lines of Lemma 2.6.

Lemma 3.4. *Let $f\in\mathcal{C}_e$. Then $|a_6|\leq 587/10800\approx 0.0543519$ and $|a_7|\leq 0.0343723$.*

Proof. A suitable rearrangement of the terms given in equation (3.3) provides us

$$345600a_6 = 5760p_5 - 480p_2p_3 + 720p_1p_4 - 480p_1p_2^2 - 17p_1^5 + 220p_1^3p_2 - 480p_1^2p_3.$$

Further, through the triangle inequality, it can be viewed as

$$\begin{aligned} 345600|a_6| \leq |5760p_5 - 480p_2p_3| + |p_1(720p_4 - 480p_2^2)| + |-17p_1^5| \\ + |p_1^2(220p_1p_2 - 480p_3)|. \end{aligned} \tag{3.36}$$

Using Lemma 2.3, we arrive at

$$|5760p_5 - 480p_2p_3| \leq 11520, \quad |p_1(720p_4 - 480p_2^2)| \leq 2880, \tag{3.37}$$

$$|p_1^2(220p_1p_2 - 480p_3)| \leq 3840, \quad \text{and} \quad |17p_1^5| \leq 544. \tag{3.38}$$

By using equation (3.37) and (3.38) in (3.36), we have

$$|a_6| \leq \frac{587}{10800} \approx 0.0543519.$$

Similarly, considering equation (3.4), we have

$$\begin{aligned} 58060800a_7 = 881p_1^6 - 13260p_1^4p_2 + 48240p_1^2p_2^2 - 106560p_1p_2p_3 + 29040p_1^3p_3 - 57600p_3^2 \\ + 69120p_1p_5 - 56160p_1^2p_4 - 86400p_2p_4 - 14400p_2^3. \end{aligned}$$

It can also be seen as with the aid of the triangle inequality,

$$\begin{aligned} 58060800|a_7| \leq |p_1^4(881p_1^2 - 13260p_2)| + |p_1p_2(48240p_1p_2 - 106560p_3)| + |p_3(29040p_1^3 - 57600p_3)| \\ + |p_1(69120p_5 - 56160p_1p_4)| + |-p_2(86400p_4 + 14400p_2^2)|. \end{aligned} \tag{3.39}$$

Lemma 2.3 takes us at

$$|p_1^4(881p_1^2 - 13260p_2)| \leq 424320, \quad |p_1p_2(48240p_1p_2 - 106560p_3)| \leq 852480, \tag{3.40}$$

$$|p_3(29040p_1^3 - 57600p_3)| \leq 921600\sqrt{\frac{15}{119}}, \quad |p_1(69120p_5 - 56160p_1p_4)| \leq 276480, \tag{3.41}$$

and

$$|p_2(86400p_4 + 14400p_2^2)| \leq 460800. \tag{3.42}$$

By substituting the values from equations (3.40)-(3.42) in (3.39), we have

$$|a_7| \leq \frac{2014080 + 921600\sqrt{\frac{15}{119}}}{58060800} \approx 0.0403246.$$

∎

We obtain the following result by omitting the proof as it directly follows from Theorem 3.1, Remark 3.2, Remark 3.3, Lemma 3.4 and equation (3.12).

Theorem 3.5. *Let $f \in \mathcal{C}_e$. Then*

$$|H_{4,1}(f)| \leq 0.00101775.$$

Acknowledgment. Neha is thankful to the Department of Applied Mathematics, Delhi Technological University, New Delhi-110042 for providing Research Fellowship.

References

[1] N. M. Alarifi, R. M. Ali and V. Ravichandran, On the second Hankel determinant for the kth-root transform of analytic functions, Filomat **31** (2017), no. 2, 227–245

[2] N. E. Cho, V. Kumar and V. Ravichandran, A survey on coefficient estimates for Carathéodory functions, Appl. Math. E-Notes **19** (2019), 370–396

[3] A. W. Goodman, *Univalent functions*. Vol. I, Mariner Publishing Co., Inc., Tampa, FL, 1983

[4] W. Janowski, Extremal problems for a family of functions with positive real part and for some related families, Ann. Polon. Math. **23** (1970/71), 159–177

[5] B. Kowalczyk, A. Lecko and Y. J. Sim, The sharp bound for the Hankel determinant of the third kind for convex functions, Bull. Aust. Math. Soc. **97** (2018), no. 3, 435–445

[6] B. Kowalczyk and A. Lecko, The sharp bound of the third Hankel determinant for functions of bounded turning, Bol. Soc. Mat. Mex. (3) **27** (2021), no. 3, Paper No. 69, 13 pp.

[7] B. Kowalczyk, A. Lecko, and D. K. Thomas, The sharp bound of the third Hankel determinant for starlike functions, Forum Mathematicum. De Gruyter, (2022)

[8] D. V. Krishna and T. RamReddy, Second Hankel determinant for the class of Bazilevic functions, Stud. Univ. Babeş-Bolyai Math. **60** (2015), no. 3, 413–420

[9] V. Kumar, N. E. Cho, V. Ravichandran and H. M. Srivastava, Sharp coefficient bounds for starlike functions associated with the Bell numbers, Math. Slovaca **69** (2019), no. 5, 1053–1064

[10] O. S. Kwon, A. Lecko and Y. J. Sim, On the fourth coefficient of functions in the Carathéodory class, Comput. Methods Funct. Theory **18** (2018), no. 2, 307–314

[11] O. S. Kwon, A. Lecko and Y. J. Sim, The bound of the Hankel determinant of the third kind for starlike functions, Bull. Malays. Math. Sci. Soc. **42** (2019), no. 2, 767–780

[12] A. Lecko, Y. J. Sim and B. Śmiarowska, The sharp bound of the Hankel determinant of the third kind for starlike functions of order 1/2, Complex Anal. Oper. Theory **13** (2019), no. 5, 2231–2238

[13] R. J. Libera and E. J. Złotkiewicz, Early coefficients of the inverse of a regular convex function, Proc. Amer. Math. Soc. **85** (1982), no. 2, 225–230

[14] W. C. Ma and D. Minda, A unified treatment of some special classes of univalent functions, in *Proceedings of the Conference on Complex Analysis (Tianjin, 1992)*, 157–169, Conf. Proc. Lecture Notes Anal., I, Int. Press, Cambridge, MA

[15] R. Mendiratta, S. Nagpal and V. Ravichandran, On a subclass of strongly starlike functions associated with exponential function, Bull. Malays. Math. Sci. Soc. **38** (2015), no. 1, 365–386

[16] A. Naz, S. Nagpal and V. Ravichandran, Starlikeness associated with the exponential function, Turkish J. Math. **43** (2019), no. 3, 1353–1371

[17] C. Pommerenke, "On the coefficients and Hankel determinants of univalent functions," J. London Math. Soc., vol. 41, pp. 111–122, 1966, doi: 10.1112/jlms/s1-41.1.111. [Online]. Available: https://doi.org/10.1112/jlms/s1-41.1.111

[18] B. Rath, K. S. Kumar, D. V. Krishna and A. Lecko, The sharp bound of the third Hankel determinant for starlike functions of order 1/2, Complex Anal. Oper. Theory **16** (2022), no. 5, Paper No. 65, 8 pp.

[19] V. Ravichandran and S. Verma, Bound for the fifth coefficient of certain starlike functions, C. R. Math. Acad. Sci. Paris **353** (2015), no. 6, 505–510

[20] J. Sokół and J. Stankiewicz, Radius of convexity of some subclasses of strongly starlike functions, Zeszyty Nauk. Politech. Rzeszowskiej Mat. No. 19 (1996), 101–105

[21] P. Zaprawa, Third Hankel determinants for subclasses of univalent functions, Mediterr. J. Math. **14** (2017), no. 1, Paper No. 19, 10 pp.

[22] P. Zaprawa, Hankel determinants for univalent functions related to the exponential function, Symm. **11** (2019), no. 3, Paper No. 10, 1211 pp.

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India
Email address: spkumar@dce.ac.in

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India
Email address: nehaverma1480@gmail.com

2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT) | 978-1-6654-1005-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICICICT54557.2022.9917683

Character Segmentation on Indian Vehicle License Plate using K-means, OTSU and Niblack algorithm

Abhishek Jha
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
abhishekjha012001@gmail.com

Aman Singh
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
aman66778@gmail.com

Abdul Basit Ali
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
basitalias11@gmail.com

Rajni Jindal
Professor
Department of Computer Science & Engineering
Delhi Technological University Delhi, India
rajnijindal@dce.ac.in

***Abstract*—Automatic License Plate Recognition (ALPR) is a growing research area as with the increase in the number of vehicles, requirements for high accuracy ALPR systems are increasing as it has been extensively used for automatic road tax collection, monitor unlike activities etc. Character segmentation is a major step in the ALPR system. This paper includes the use of image pre-processing techniques, K-means clustering, Otsu and Niblack method and Connected Component Analysis (CCA) to perform character segmentation. Unlike existing works which are based on license plates of other countries and therefore may not perform well on Indian license plates, in this work we have dealt with Indian license plates. Many of the existing works are based on license plates that are captured from a straight angle and good illumination and hence may not perform well if the images of license plates have noises or are captured at different angles. Therefore, in this work we have used license plates which are at different angles with different lighting conditions.**

Keywords—Character Segmentation, Binarization, Connected Component Analysis (CCA), Indian License Plates

I. INTRODUCTION

In modern times it has become quite necessary to be able to monitor and analyze the traffic on the road to help maintain the integrity and safety of transport facilities. ALPR is a solution that can take care of the monitoring part of the problem. ALPR can be implemented using existing CCTV surveillance cameras or we can also have dedicated cameras for Number plate detection. Systems like this use optical character recognition (OCR) software to segment and identify vehicle number plate details. This technology is used to detect speeding violations, toll collection, or to detect any criminal movement, but it can also be used for access control and to monitor any vehicle movement.

ALPR systems have been under development since the 1970s and have become very reliable and cost effective. A typical ALPR system includes software and hardware components including roadside camera systems, control computer systems, applications to manage data, and a central database of vehicle number plate details.

ALPR system contains four segments:

Vehicle detection: Vehicle detection is the first step towards license plate recognition. Vehicle is localized in the image and defines the area of interest for further license plate recognition process. YOLO (you only look once) object detection models is used by many researchers for the purpose of vehicle detection [1]. Other deep learning like CNN (convolutional neural networks) are also used.

License plate detection: License plate is localized on the vehicle in the vehicle image detected from vehicle detection segment in this segment. There are many methods for this segment like edge based, Color-based deep learning models etc. In one of the paper based on edge based method, only vertical edges have been used to detect the license plate [2]. Some used the fact of the difference in the color of license plate to the color of vehicle in the color-based method.

Character segmentation: Characters are localized from the license plates. Deep learning, image processing techniques such as projection methods and pixel connectivity method are used for character segmentation. In the projection method vertical and horizontal projections are used to extract the character [3]. CNN (Convolutional Neural Networks) are also used to create bounding boxes around the character of license plates in deep learning methods [4] but they require large datasets for training as compared to other methods.

Character recognition: Character recognition is the final step of the ALPR system that is to identify the characters of the license plate. It can be achieved by various methods such as pattern matching, deep learning, template matching [5].

In our paper we focus on the character segmentation part of ALPR. We have used methods like OTSU [6], K-means clustering [7] and Niblack method [8] to binarize the data, i.e., to divide the image of a number plate into two sets in order to differentiate between plate background and the character segments. After the binarization of the number plate, we use connected component analysis to detect character boundaries within the number plate.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:29:07 UTC from IEEE Xplore. Restrictions apply.

II. RELATED WORK

Character segmentation is the determining factor of an ALPR system that has the most effect on accuracy of the system. Hence, a lot of work has been done to improve the accuracy of the character segmentation of license plates. The character segmentation of license plates considers the attribute of license plates of having contrasting colours for the characters and the background. The binarization of the license plates make the separation of the characters and the background easier for further work.

There are many local thresholding binarization methods out of which some popular ones are briefly described below. In local thresholding, a threshold for each pixel is calculated by inspecting the surrounding area of that pixel. In the method proposed by Sauvola et al. [9], they proposed a formula for computing the pixel-wise thresholds. They calculated the pixel-wise threshold using the factors m and s which are the mean and the standard deviation of the surrounding region which is of square window of width N, k which is a weight variable controlling the level of threshold and R which is the dynamic range of the standard deviation.

In global thresholding methods ,there is a single threshold value for the entire image. But these methods have issues when the image has shadows or uneven illumination. The Otsu method [6] is one of the most popular global thresholding method. It automatically returns the global threshold for the intensity that divides pixels of the image to two groups. Kittler and Illingworth [9] proposed the method in which global thresholding is computed by modelling the foreground and background into two Gaussian distributions. Multimethod binarization method has also been proposed which tries to utilize the properties of binarized license plates images produced by the various binarization methods. Y Yoon et al. [10] in his paper, used this method and showed combinations of four different binarization methods showed better results than single method binarization.

There are many existing methods for segmentation by considering their advantages and limitations, some of which are described below.

In pixel connectivity method, connected pixels are labelled and if they are of the same label for a predetermined size or aspect ratio, they are extracted as a potential character of the license plate. This method is used by Wu, Bing-Fei & Lin in this paper [11]. The advantage of this method is that it is a simple and effective method and usually takes less computation power and time. It is also less affected by rotation of license plates. However, the drawback with this method is that it has issues if characters are joined or broken due to the binarization thresholding method. Therefore, in this method one of the determining factors is the binarization thresholding method.

The projection profiles method for character segmentation is based on the fact that the characters and the background have different colours after binarization. The starting and ending positions of the character are detected using the vertical projections while horizontal projections are used to extract the characters [3]. These methods give low robust values when compared to pixel connectivity methods if the images are rotated, yet they are still robust to rotations. One issue with these methods is that they are sensitive to image quality and noise.

Some papers have used prior knowledge about the license plates such as the aspect ratio, the ratio of various coloured pixels, the typical positions of the characters for character segmentation. Paliy et al. [12] used a simple approach by resizing the license plate to a specific dimension, where they had predetermined the character positions. These methods are simple to implement and give results usually specific to the regions where they are designed to operate and hence may not give good results if generalized in other regions.

With the growth in computation power of machines, deep learning is now giving great results in many fields. Neural networks have also been used for character segmentation [4]. Typically, in these papers Convolutional Neural Networks (CNN) are used for the task associated with computer vision. A license plate is given as the input to the neural network and the output generally is the bounding boxes of the characters of the license plates. In some papers, several deep learning based license plate recognition pipelines have reduced the number of parameters of the network by omitting character segmentation explicitly in place of implicit character segmentation in later stages. However, these methods require more resources, time and dataset as compared to other methods described above.

III. DATASET

The dataset consisting of Indian license plate images are used for this work. Two datasets have been used: Car License Plate dataset[13] contains 433 images and Indian Number Plate Detection[14] dataset contains 373 images in total.

49 images are selected from 433 images of Car License Plate dataset and 235 images are selected from the 373 images of Indian Number Plate Detection dataset gives a total of 284 images based on the Indian origin, human readable clarity, removing duplicate images(a lot of images have many duplicate copies). In these given datasets, initially we are provided with the car images out of which, we need to extract license plates. In the dataset Car License Plate dataset[13], annotations of number plates are already provided which can be used for extracting license plates whereas in the dataset Indian Number Plate Detection[14], no such annotations are provided, hence we need to manually annotate the license plates and then extract them using these annotations.

After this step, we need to annotate the characters in the number plates of both dataset which has been done manually. All the annotations performed have been done using the LabelImg tool [15] which is a graphical image annotation tool. Some sample images before processing are shown in Fig. 1.



Fig.1 Sample Images of the dataset before processing

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:29:07 UTC from IEEE Xplore. Restrictions apply.

IV. PROPOSED WORK

In this section the work done in order to detect the character from the Indian license plate is discussed. In the first step, the license plates are extracted from the car images using the annotations as shown in Fig. 2(a).

A. Image Preprocessing

1) Grayscaling: RGB Images are first converted to the grayscale image. Grayscale image consists of pixels of intensity in the range of [0,255] where 0 intensity pixel is a dark pixel and 255 intensity pixel is bright pixel and the pixels between 0 and 255 are shades of gray. License plate image after grayscaling the original extracted license plate is shown in Fig. 2(b).

2) Normalization: Min Max normalization is performed to ensure similar distribution of intensity of image pixels. Min max normalization can be performed by (1).

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

where x is the original value of pixel while x' is normalized value of the pixel.

3) Image Sharpening: Components that should be classified as two different components can be classified as one component due to a very fine gap between them. Sharpness of image is helpful in separating out these components by enhancing the intensity of the fine gap between them. Sharpened image is obtained by convolving the below kernel as in (2) on the image [16]. License plate image after performing image sharpening is shown in Fig. 2(c).

$$\text{Image sharpening kernel} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad (2)$$

B. Image Binarization

Image binarization refers to converting the grayscale image to the Black and white image. Here the image pixels are classified in two groups (black and white) and each pixel belongs to exactly one of the groups (either the black pixels group or the white pixels group). Image binarization can be performed in many different ways.

1) K-means: K-means [7] clustering algorithm assigns the n number of observations to the k clusters. Here the k-means clustering algorithm is used to assign each pixel of the license plate image to the 2 (k = 2) cluster that is black or white cluster. Binary image of license plate done after applying K-means clustering method is shown in Fig. 3(a).

2) Otsu: Otsu method [6] automatically returns the global threshold for the intensity that divides pixels of the image to two groups. Otsu searches for the global threshold intensity(t) that minimizes the internal variance of the pixels in both groups σ_0^2 & σ_1^2 as shown in (4). Binary image of license plate after applying Otsu method is shown in Fig. 3(b).

$$\sigma_\omega^2(t) = \omega_0(t)\,\sigma_0^2(t) + \omega_1(t)\,\sigma_1^2(t) \quad (4)$$

ω_0 and ω_1 are probabilities of the two groups divided by threshold t.

3) Niblack: Niblack method is one of the popular local thresholding methods that calculates the threshold for each pixel by inspecting the surrounding region of that pixel. For each pixel, a surrounding region is a square window of width N. Binary image of license plate after applying Niblack method is shown in Fig. 3(c).



(b) Original Extracted License Plate



(b) Grayscale image



(c) Sharpened image

Fig. 2 Image Preprocessing



(a) K-Means



(b) OTSU Method



(c) Niblack Method

Fig. 3 Image Binarization

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:29:07 UTC from IEEE Xplore. Restrictions apply.

In Niblack method, local threshold for each pixel is computed using the given formula :

$$T = m + k * s$$

- m is the mean of the surrounding region of the pixel for which threshold is computed
- s is the standard deviation of the surrounding region of the pixel for which threshold is computed
- k is weight variable that controls the level of thresholds.

Here, both N and k can be adjusted depending on the images.

C. Connected Component Analysis

Connected component analysis is a known technique that forms the component of the pixels based upon the pixel connectivity. Two pixels that are of the same group (either of black or white) and are neighbors of each other are said to be connected and are the part of the same connected component. Here we have considered the 8 neighbors method. A pixel is a neighbor of 4 pixels sharing the side and also the neighbor of 4 pixels that is sharing the corner when represented in the form of a two-dimensional matrix.

Components are rejected based on the following basis:

- Minimum and maximum ratio of height of the component to the height of the License Plate.
- Minimum and maximum ratio of the width of the component to the width of the License Plate.
- Range of the height to width ratio of the component.
- Range of ratio of area of component to the area of License Plate.

Top 10 components based on the area are selected as the detected character of a given license plate.

D. Evaluation Metrics

IoU (Intersection over union) is simple and effective metric for evaluating or measuring the overlap between two bounding boxes. IoU is defined (5).

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (5)$$

A threshold is used for the intersection over union i.e., if the IoU value for two bounding boxes is greater than or equal to the fixed threshold value then two boxes are said to be overlapped.

V. EXPERIMENT & RESULTS

The performance of the character segmentation process with these three algorithms K-means, Otsu and Niblack on Indian License Plate images are shown in Table 1. K-means clustering algorithm is an iterative algorithm. We fixed the maximum iteration to be 10, maximum iteration can be increased to get more precision over image binarization for the K-means.

In the Niblack method, we have used k to be 0.2 and N to be 15.

Threshold value for the IoU is fixed to 0.4 for this experiment. threshold value can be altered according to the dataset.

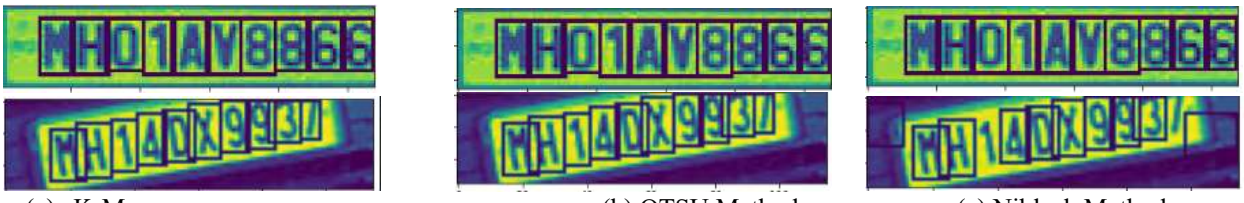
VI. CONCLUSION & FUTURE WORK

The performance of character detection of license plates by these methods do not get affected by rotation of images i.e. performance does not depend on the orientation of the characters in the license plate. The resources and time required by these methods are lower than resources and time required if deep learning methods are used. There is also no requirement of a large dataset for training as these methods do not need to be trained to give results.

Further, as an extension to the above work, to get better performance multimethod binarization could be used so that all the methods shown above could be used in a single method so that the advantages of each of the methods could be utilized in a single method.

TABLE I. CHARACTER SEGMENTATION RESULTS

METHOD	TRUE POSITIVE	FALSE POSITIVE	FALSE NEGATIVE	ACCURACY
K-MEANS	2521	90	221	89.01%
OTSU	2523	88	219	89.15%
NIBLACK	2516	213	226	85.14%



(a) K-Means (b) OTSU Method (c) Niblack Method

Fig. 4 Character Segmentation

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:29:07 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] Wang, Y.-N., Pang, Z.-H., & Yuan, D.-M. (2016). Vehicle Detection Based on YOLO in Real Time. *Wuhan Ligong Daxue Xuebao/Journal of Wuhan University of Technology*, *38*, 41–46. https://doi.org/10.3963/j.issn.1671-4431.2016.10.008.

[2] Sarfraz, M., Ahmed, M. J., & Ghazi, S. A. (2003). Saudi Arabian license plate recognition system. *2003 International Conference on Geometric Modeling and Graphics, 2003. Proceedings*, 36–41. https://doi.org/10.1109/GMAG.2003.1219663.

[3] Du, S., Ibrahim, M., Shehata, M., & Badawy, W. (2013). Automatic License Plate Recognition (ALPR): A State-of-the-Art Review. *IEEE Transactions on Circuits and Systems for Video Technology*, *23*(2), 311–325. https://doi.org/10.1109/TCSVT.2012.2203741.

[4] Montazzolli, S., & Jung, C. (2017). Real-Time Brazilian License Plate Detection and Recognition Using Deep Convolutional Neural Networks. *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 55–62. https://doi.org/10.1109/SIBGRAPI.2017.14.

[5] Rahman, C. A., Badawy, W., & Radmanesh, A. (2003). A real time vehicle's license plate recognition system. *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, 2003.*, 163–166. https://doi.org/10.1109/AVSS.2003.1217917.

[6] Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076.

[7] Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 881–892. https://doi.org/10.1109/TPAMI.2002.1017616.

[8] W. Niblack, *An Introduction to Image Processing, Englewood Cliffs*, NJ: Prentice-Hall, 1986, pp. 115-116.

[9] Kittler, J., & Illingworth, J. (1986). Minimum error thresholding. *Pattern Recognition*, *19*(1), 41–47. https://doi.org/https://doi.org/10.1016/0031-3203(86)90030-0.

[10] Yoon, Y., Ban, K.-D., Yoon, H., Lee, J., & Kim, J. (2013). Best Combination of Binarization Methods for License Plate Character Segmentation. *ETRI Journal*, *35*(3), 491–500. https://doi.org/https://doi.org/10.4218/etrij.13.0112.0545.

[11] Wu, B.-F., Lin, S.-P., & Chiu, C.-C. (2007). Extracting characters from real vehicle licence plates out-of-doors. *Computer Vision, IET*, *1*, 2–10. https://doi.org/10.1049/iet-cvi:20050132.

[12] Paliy, I., Turchenko, V., Koval, V., Sachenko, A., & Markowsky, G. (2004). *Approach to recognition of license plate numbers using neural networks*. *4*, 2965–2970 vol.4. https://doi.org/10.1109/IJCNN.2004.1381137.

[13] Car License Plates Dataset. (n.d.). In *Make ML*. https://makeml.app/datasets/cars-license-plates.

[14] C. Kumar, "Indian number plate detection," *Kaggle*, 29-Jul-2020. [Online]. Available: https://www.kaggle.com/datasets/ckay16/indian-number-plate-detection. [Accessed: 26-Apr-2022].

[15] P. Saxena, "How to label images for object detection, step by step," *Medium*, 12-Aug-2021. [Online]. Available: https://towardsdatascience.com/how-to-label-images-for-object-detection-step-by-step-7ee317f98583. [Accessed: 26-Apr-2022].

[16] "Sharpening an image using opencv library in Python," *Analytics Vidhya*, 26-Aug-2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/08/sharpening-an-image-using-opencv-library-in-python/. [Accessed: 26-Apr-2022].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:29:07 UTC from IEEE Xplore. Restrictions apply.

arXiv:2210.13166v1 [math.CV] 24 Oct 2022

Coefficient Functional and Bohr-Rogosinski Phenomenon for Analytic functions involving Semigroup Generators

Surya Giri and S. Sivaprasad Kumar

Abstract

This paper examines the coefficient problems for the class of semigroup generators, a topic in complex dynamics that has recently been studied in context of geometric function theory. Further, sharp bounds of coefficient functional such as second order Hankel determinant, third order Toeplitz and Hermitian-Toeplitz determinants are derived. Additionally, the sharp growth estimates and the bounds of difference of successive coefficients are determined, which are used to prove the Bohr and the Bohr- Rogosinski phenomenon for the class of semigroup generators.

Keywords: Holomorphic generators; Hankel determinant; Toeplitz determinant; Zalcman functional; Successive coefficient difference; Bohr and Bohr-Rogosinski radius.
AMS Subject Classification: 30C45, 30C50, 30C55, 47H20, 37L05.

1 Introduction

Let $\mathcal{H}$ be the class of holomorphic functions in the unit disk $\mathbb{D}$ and $A \subset \mathcal{H}$ containing functions of the form

$$f(z) = z + \sum_{n=2}^{\infty} a_n z^n. \tag{1.1}$$

By $\mathcal{B}$, we denote the class of holomorphic self mappings from $\mathbb{D}$ to $\mathbb{D}$. A family $\{u_t(z)\}_{t\geq 0} \subset \mathcal{B}$ is called a one parameter continuous semigroup if (i) $\lim_{t\to 0} u_t(z) = z$, (ii)$u_{t+s}(z) = u_t(z) \circ u_s(z)$, and (iii) $\lim_{t\to s} u_t(z) = u_s(z)$ for each $z \in \mathbb{D}$ hold.

Berkson and Porta [3] showed that each one parameter semigroup is locally differentiable in parameter $t \geq 0$ and moreover, if

$$\lim_{t\to 0} \frac{z - u_t(z)}{t} = f(z),$$

which is a holomorphic function, then $u_t(z)$ is the solution of the the Cauchy problem

$$\frac{\partial u_t(z)}{\partial t} + f(u_t(z)) = 0, \quad u_0(z) = z.$$

The function f is called the infinitesimal (holomorphic) generator of semigroup $\{u_t(z)\} \subset \mathcal{B}$. The class of all holomorphic generators is denoted by $\mathcal{G}$. Also, note that each element of $\{u_t(z)\}$ generated by $f \in \mathcal{H}(\mathbb{D}, \mathbb{C})$ is univalent function while f is not necessarily univalent [13]. Different analytic criteria are available to show that a function is semigroup generator [3, 37, 6]. Berkson and Porta [3] proved:

Theorem 1.1. *The following assertions are equivalent:*

(*a*) $f \in \mathcal{G}$;

(*b*) $f(z) = (z - \sigma)(1 - z\bar{\sigma})p(z)$ *with some* $\sigma \in \overline{\mathbb{D}}$ *and* $p \in \mathcal{H}$, $\mathrm{Re}(p(z)) \geq 0$.

The point $\sigma \in \overline{\mathbb{D}} := \{z \in \mathbb{C} : |z| \leq 1\}$ is unique and called the Denjoy–Wolff point of the semigroup generated by $f(z)$. By Denjoy-Wolff theorem [13, 37, 36] for continuous semigroup, $\lim_{t\to\infty} u(t,z) = \sigma$ if the semigroup generated by f is neither an elliptic automorphism of $\mathbb{D}$ nor an identity map for at least one $t \in [0,\infty)$. We denote the class of infinitesimal generators with Denjoy-Wolff point σ by $\mathcal{G}[\sigma]$. For $\sigma = 0$, we obtain the following subclass

$$\mathcal{G}[0] = \{f \in \mathcal{G} : f(z) = zp(z), \ \ \operatorname{Re} p(z) \geq 0\}.$$

Bracci et al. [7] considered the class $\mathcal{G}_0 = \mathcal{G}[0] \cap \mathcal{A}$. In the study of non-autonomous problem such as Loewner theory, the class $\mathcal{G}_0$ plays a significant role [8, 12]. Various subclasses of $\mathcal{G}_0$ are considered with parameter, which is also called filtration. For instance, the class

$$\mathcal{A}_\beta = \left\{ f \in \mathcal{A} : \operatorname{Re}\left(\beta \frac{f(z)}{z} + (1-\beta) f'(z)\right) > 0, \quad \beta \in [0,1] \right\} \tag{1.2}$$

is a subclass of $\mathcal{G}_0$. In [7], the authors proved that $\mathcal{A}_{\beta_1} \subsetneq \mathcal{A}_{\beta_2} \subsetneq \mathcal{G}_0$ for $0 \leq \beta_1 < \beta_2 < 1$ and whenever $f \in \mathcal{A}_\beta$,

$$\operatorname{Re} \frac{f(z)}{z} \geq \int_0^1 \frac{1 - t^{1-\beta}}{1 + t^{1-\beta}} dt.$$

Clearly, when $\beta = 0$, the class $\mathcal{A}_\beta$ reduces to the class $\mathcal{R} := \{f \in \mathcal{A} : \operatorname{Re} f'(z) > 0\}$, which is called the class of bounded turning functions. It can be easily seen that $\mathcal{R} \subset \mathcal{G}_0$ and each $f \in \mathcal{R}$ satisfies the Noshiro-Warshawski condition [12], thus every member of $\mathcal{R}$ generates a semigroup that is univalent. Elin et al. [14] solved the radii problems for the class $\mathcal{A}_\beta$. They found the radii $r \in (0,1)$ for $f \in \mathcal{A}_\beta$ such that $f(rz)/r$ belong to the class of starlike functions, denoted by $\mathcal{S}^*$, and some other subclasses of starlike functions. This problem arises from the fact that neither $\mathcal{S}^* \subset \mathcal{A}_\beta$ nor $\mathcal{A}_\beta \subset \mathcal{S}^*$. Generalizing this work, Giri and Kumar [16] obtained r such that $f(rz)/r$ belong to the class of Ma-Minda starlike functions.

Coefficient problems, growth estimates and others were still open for the class $\mathcal{A}_\beta$. In this paper, we focus on these problems. We find the bound of n^{th} coefficient of $f \in \mathcal{A}_\beta$ and coefficient functional such as second Hankel determinant, third order Toeplitz and Hermitian Toeplitz determinant and Zalcman functional. Later, Bohr and Bohr-Rogosinski phenomenon with growth estimates are also discussed for this class.

In 1914, Bohr [5] proved that, if $\omega(z) = \sum_{n=0}^{\infty} c_n z^n \in \mathcal{B}$, then $\sum_{n=0}^{\infty} |c_n| r^n \leq 1$ for all $z \in \mathbb{D}$ with $|z| = r \leq 1/3$. The constant $1/3$ is known as Bohr radius. This inequality was first derived by Bohr for $r \leq 1/6$ and sharpened independently to $r \leq 1/3$ by Wiener, Riesz, and Schur. In recent years, a lot of study have been carried out on the Bohr inequality for functions, which map $\mathbb{D}$ onto other domains, say Ω [4, 15]. Different generalizations of the Bohr inequality are taken into consideration [39, 27]. We say that

Definition 1.2. The class $\mathcal{A}_\beta$ satisfies the Bohr phenomenon if there exists r_b such that

$$|z| + \sum_{n=2}^{\infty} |a_n| |z|^n \leq d(f(0), \partial f(\Omega))$$

holds in $|z| = r \leq r_b$, where $\partial f(\Omega)$ is the boundary of image domain of $\mathbb{D}$ under f and d denotes the Euclidean distance between $f(0)$ and $\partial f(\Omega)$.

For $m = 1$, Muhanna [32] showed that the Bohr phenomenon holds for the class of univalent functions and the class of convex functions, when $|z| = r \leq 3 - 2\sqrt{2}$ and $|z| = r \leq 1/3$ respectively. We refer to the survey article [31] for further details on this topic. There is also the concept of Rogosinski radius along with the Bohr radius, although a little is known about Rogosinski radius in comparison to Bohr radius [17, 22, 35]. It says that, if $\omega(z) = \sum_{n=0}^{\infty} c_n z^n \in \mathcal{B}$, then

$$\sum_{n=0}^{N-1} |c_n| |z|^n \leq 1 \quad (N \in \mathbb{N})$$

in the disk $|z| = r \leq 1/2$. The radius $1/2$ is called the Rogosinski radius. Kayumov et al. [18] considered the following expression

$$R_N^f(z) := |f(z)| + \sum_{n=N}^{\infty} |a_n| |z|^n$$

and found the radius r_N such that $R_N^f(z) \leq 1$ in $|z| = r \leq r_N$ for the Cesáro operators on the space of bounded analytic functions. Here, we say that:

Definition 1.3. The class $\mathcal{A}_\beta$ satisfies the Bohr-Rogosinski phenomenon if there exist r_N such that

$$|f(z^m)| + \sum_{n=N}^{\infty} |a_n||z|^n \leq d(f(0), \partial f(\Omega))$$

holds in $|z| = r \leq r_N$.

Section 5 is devoted to find the r_b and r_N for the class $\mathcal{A}_\beta$.

For $f(z) = z + \sum_{n=2}^{\infty} a_n z^n \in \mathcal{A}$, the m^{th} Hankel, Toeplitz and Hermitian Toeplitz determinant for $m \geq 1$ and $n \geq 0$ are respectively given by

$$H_m(n)(f) = \begin{vmatrix} a_n & a_{n+1} & \cdots & a_{n+m-1} \\ a_{n+1} & a_{n+2} & \cdots & a_{n+m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n+m-1} & a_{n+m} & \cdots & a_{n+2m-2} \end{vmatrix},$$

$$T_m(n)(f) = \begin{vmatrix} a_n & a_{n+1} & \cdots & a_{n+m-1} \\ a_{n+1} & a_n & \cdots & a_{n+m-2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n+m-1} & a_{n+m-2} & \cdots & a_n \end{vmatrix}, \tag{1.3}$$

$$T_{m,n}(f) = \begin{vmatrix} a_n & a_{n+1} & \cdots & a_{n+m-1} \\ \bar{a}_{n+1} & a_n & \cdots & a_{n+m-2} \\ \vdots & \vdots & \vdots & \vdots \\ \bar{a}_{n+m-1} & \bar{a}_{n+m-2} & \cdots & a_n \end{vmatrix}, \tag{1.4}$$

where $\bar{a}_n = \overline{a_n}$. Toeplitz matrices have constant entries along their diagonals, while Hankel matrices have constant entries along their reverse diagonals. In particular,

$$H_2(n)(f) = a_n a_{n+2} - a_{n+1}^2, \;\; T_3(1)(f) = 1 - 2a_2^2 + 2a_2^2 a_3 - a_3^2$$

and $T_{3,1}(f) = 1 - 2|a_2|^2 + 2\,\mathrm{Re}(a_2^2 \bar{a}_3) - |a_3|^2$. Finding the sharp bound of $|H_2(2)(f)|$ for the class $\mathcal{S}$ and its subclasses has always been the focus of many researchers. Although, investigations concerning Toeplitz and Hermitian Toeplitz are recently introduced in [2, 11], a summary of some of the more significant results is given in [38]. For more work in this direction (see [21, 20, 24, 33]).

In 1999, Ma [28] proposed a conjecture for $f(z) = z + \sum_{n=2}^{\infty} a_n z^n \in \mathcal{S}$ that

$$|J_{m,n}| := |a_n a_m - a_{n+m-1}| \leq (m-1)(n-1).$$

They proved this conjecture for the class of starlike functions and univalent functions with real coefficients. It is also called generalized Zalcman conjecture as it generalizes the Zalcman conjecture $|a_n^2 - a_{2n-1}| \leq (2n-1)^2$ for $f \in \mathcal{S}$. Recently, bound of $|J_{2,3}|$ are obtained for various subclasses of $\mathcal{A}$ [1, 10]. In section 2 and 3, we obtain the sharp bound of $|H_2(2)(f)|$, $|T_3(1)(f)|$ and $|J_{2,3}(f)|$ for $f \in \mathcal{A}_\beta$.

2 Hankel Determinant and Zalcman Functional

Theorem 2.1. *If $f \in \mathcal{A}_\beta$ is of the form (1.1), then*

$$|a_n| \leq \frac{2}{n - \beta(n-1)}. \tag{2.1}$$

Further, this inequality is sharp for each n.

Proof. Let $f \in \mathcal{A}_\beta$ is given by (1.1), then we have

$$\beta\frac{f(z)}{z} + (1-\beta)zf'(z) = p(z) \qquad (z \in \mathbb{D}),$$

where $p(z) = 1 + \sum_{n=1}^{\infty} p_n z^n$ such that $\operatorname{Re} p(z) > 0$ is a member of the Carathéodory class $\mathcal{P}$. Upon comparing the coefficients of same powers on either side with the series expansion of f and p yields

$$(n-(n-1)\beta)a_n = p_{n-1} \tag{2.2}$$

for $n = 2, 3, 4, \cdots$, which gives the needed bound of $|a_n|$ using the Carathéodory coefficient bounds $|p_n| \leq 2$ (see [12]). The function $\tilde{f} : \mathbb{D} \to \mathbb{C}$ defined by

$$\tilde{f}(z) = z\bigg(-1 + 2\bigg({}_2F_1\bigg[1, \frac{1}{1-\beta}, \frac{2-\beta}{1-\beta}, z\bigg]\bigg)\bigg) = z + \sum_{n=2}^{\infty} \frac{2}{n-(n-1)\beta} z^n \tag{2.3}$$

satisfies the condition $\operatorname{Re}\Big(\beta\tilde{f}(z)/z + (1-\beta)\tilde{f}'(z)\Big) > 0$, hence $\tilde{f}$ is a member of $\mathcal{A}_\beta$, where ${}_2F_1$ denotes the Gauss hypergeometric function. Equality in (2.1) occurs for $\tilde{f}$, which proves the sharpness of the bound.

Corollary 2.2. *If $f \in \mathcal{A}_\beta$, then for any real $\mu \geq 0$*

$$|a_n a_{n+2} - \mu a_{n+1}^2| \leq \frac{4}{(n-(n-1)\beta)(n+2-(n+1)\beta)} + \frac{4\mu}{(n+1-n\beta)^2}.$$

The bound is sharp.

Proof. Since $|a_n a_{n+2} - \mu a_{n+1}^2| \leq |a_n||a_{n+2}| + \mu|a_{n+1}|^2$. The bound simply follows from (2.1). To see the sharpness, consider

$$\tilde{f}_1(z) = z\left(-1 + 2\left({}_2F_1\left[1, \frac{1}{1-\beta}, \frac{2-\beta}{1-\beta}, iz\right]\right)\right) = z + \sum_{n=2}^{\infty} \frac{2i^{n-1}}{(n-(n-1)\beta)} z^n. \tag{2.4}$$

It can be easily seen that $\tilde{f}_1(z)$ satisfy (1.2), thus $\tilde{f}_1 \in \mathcal{A}_\beta$.

For $\mu = 1$, Corollary 2.2 gives the following sharp bound:

Corollary 2.3. *If $f \in \mathcal{A}_\beta$ is of the form (2.2), then*

$$|H_2(n)(f)| \leq \frac{4\left((2n^2-1)\beta^2 - (4n^2+4n-2)\beta + 2n^2+4n+1\right)}{(n-(n-1)\beta)(n+2-(n+1)\beta)(n+1-n\beta)^2}.$$

For $n = 2$ and 3, the following sharp bound of second order Hankel determinant follows:

Corollary 2.4. *If $f \in \mathcal{A}_\beta$ is of the form (2.2), then*

$$|H_2(2)(f)| \leq \frac{4(7\beta^2 - 22\beta + 17)}{(4-3\beta)(3-2\beta)^2(2-\beta)}, \quad |H_2(3)(f)| \leq \frac{4(17\beta^2 - 46\beta + 31)}{(5-4\beta)(4-3\beta)^2(3-2\beta)}.$$

Theorem 2.5. *If $f \in \mathcal{A}_\beta$ is of the form (1.1), then*

$$|J_{2,3}(f)| \leq \frac{2}{4-3\beta}.$$

The bound is sharp.

Proof. Let $f \in \mathcal{A}_\beta$ is given by (1.1), then from (2.2), we have

$$|J_{2,3}(f)| = |a_2 a_3 - a_4| = \left|\frac{p_1 p_2}{(3-2\beta)(2-\beta)} - \frac{p_3}{4-3\beta}\right|. \tag{2.5}$$

For $p(z)=1+\sum_{n=1}^{\infty}p_nz^n\in\mathcal{P}$, Libera et al. [26] proved that

$$\begin{aligned}2p_2&=p_1^2+x(4-p_1^2),\\4p_3&=p_1^3+2xp_1(4-p_1^2)-x^2p_1(4-p_1^2)+2z(1-|x|^2)(4-p_1^2),\end{aligned}\tag{2.6}$$

where $|x|\leq 1$ and $|z|\leq 1$. Substituting these values of p_2 and p_3 in (2.5), we obtain

$$\begin{aligned}|J_{2,3}(f)|=\bigg|\frac{p_1^3}{4}\bigg(\frac{2}{2\beta^2-7\beta+6}+\frac{1}{3\beta-4}\bigg)-\frac{p_1(4-p_1^2)(1-\beta)^2x}{(2-\beta)(3-2\beta)(4-3\beta)}\\+\frac{p_1(4-p_1^2)x^2}{4(4-3\beta)}-\frac{(4-p^2)(1-|x|^2)z}{2(4-3\beta)}\bigg|.\end{aligned}$$

Since the class $\mathcal{P}$ is rotationally invariant and it is an easy exercise to check that the class $\mathcal{A}_\beta$ is also rotationally invariant, therefore, without losing generality, we can take $p_1=p\in[0,2]$. Now, applying the triangle inequality with $|x|=\rho$, we obtain

$$\begin{aligned}|J_{2,3}(f)|\leq&\frac{p^3}{4}\bigg(\frac{2}{2\beta^2-7\beta+6}+\frac{1}{3\beta-4}\bigg)+\frac{p(4-p^2)(1-\beta)^2\rho}{(2-\beta)(3-2\beta)(4-3\beta)}+\frac{4-p^2}{2(4-3\beta)}\\&+\rho^2\bigg(\frac{p(4-p^2)}{4(4-3\beta)}-\frac{(4-p^2)}{2(4-3\beta)}\bigg)=:F(p,\rho).\end{aligned}$$

To determine the maximum value of $F(p,\rho)$, first we find out the stationary points, given by the roots of $\partial F/\partial p=0$ and $\partial F/\partial\rho=0$, where

$$\begin{aligned}\frac{\partial F(p,\rho)}{\partial p}=&\frac{3p^2(r^2(2\beta^2-7\beta+6)+4r(1-\beta)^2+2\beta^2-\beta-2)}{4(-2+\beta)(-3+2\beta)(-4+3\beta)}+p\bigg(\frac{r^2}{4-3\beta}-\frac{1}{4-3\beta}\bigg)\\&+\frac{r^2}{4-3\beta}+\frac{4r(1-\beta)^2}{(4-3\beta)(3-2\beta)(2-\beta)}.\\\frac{\partial F(p,\rho)}{\partial\rho}=&2r\bigg(\frac{p(4-p^2)}{4(4-3\beta)}+\frac{-4+p^2}{2(4-3\beta)}\bigg)+\frac{p(4-p^2)(1-\beta)^2}{(4-3\beta)(3-2\beta)(2-\beta)}.\end{aligned}$$

A simple calculation shows that for $p\in[0,2]$ and $r\in[0,1]$, the stationary point is $(0,0)$ and

$$\bigg(\frac{\partial^2F}{\partial p^2}\frac{\partial^2F}{\partial\rho^2}-\frac{\partial^2F}{\partial\rho\partial p}\bigg)_{(p,\rho)=(0,0)}=\frac{4(8-11\beta+4\beta^2)}{(3-2\beta)^2(2-\beta)^2(4-3\beta))}>0\quad\text{for all }\ \beta\in[0,1].$$

Thus $F(p,\rho)$ attains either maximum or minimum at $(p,\rho)=(0,0)$. Since, we have

$$\bigg(\frac{\partial^2F}{\partial p^2}\bigg)_{(0,0)}=\frac{-1}{4-3\beta}<0,\ \ \bigg(\frac{\partial^2F}{\partial\rho^2}\bigg)_{(0,0)}=\frac{-4}{4-3\beta}<0\quad\text{for all}\quad\beta\in[0,1].$$

Therefore, $F(p,\rho)$ attain its maximum value at $(p,\rho)=(0,0)$, which is $2/(4-3\beta)$.

Now, to prove the sharpness of the bound, consider the function $\tilde{f}_2:\mathbb{D}\to\mathbb{C}$ given by

$$\alpha\frac{\tilde{f}_2(z)}{z}+(1-\alpha)\tilde{f}_2'(z)=\frac{1+z^3}{1-z^3}.\tag{2.7}$$

If $\tilde{f}_2(z)=z+\sum_{n=2}^{\infty}a_nz^n$, then $a_2=a_3=0$ and $a_4=2/(4-3\beta)$, thus $|J_{2,3}(f)|=2/(4-3\beta)$.

3 Toeplitz and Hermitian-Toeplitz Determinant

Theorem 3.1. *If $f\in\mathcal{A}_\beta$ is of the form (1.1), then*

(i) $|T_{2,n}(f)| \leq 4\left(\dfrac{1}{(n-\beta(n-1))^2}+\dfrac{1}{(n+1-n\beta)^2}\right)$,

(ii) $|T_{3,1}(f)| \leq \dfrac{4\beta^4-28\beta^3+101\beta^2-196\beta+140}{(3-2\beta)^2(\beta-2)^2}$.

The bounds are sharp.

Proof. From (1.3), it follows that

$$|T_{2,n}(f)| = |a_n^2 - a_{n+1}^2| \leq |a_n|^2 + |a_{n+1}|^2.$$

Using the bound of $|a_n|$ from (2.1), required bound of $|T_{2,n}(f)|$ follows directly and equality case holds for the function $\tilde{f}_1$ given by (2.4).

Now we proceed for $|T_{3,1}(f)|$. Again from (1.3), we have

$$|T_{3,1}(f)| = |1-2a_2^2+2a_2^2a_3-a_3^2| \leq 1+2|a_2|^2+|a_3||a_3-2a_2^2|. \tag{3.1}$$

By (2.2),

$$|a_3-2a_2^2| = \frac{1}{3-2\beta}\left|p_2-\frac{2(3-2\beta)}{(2-\beta)^2}p_1^2\right|.$$

Applying the well known result $|p_2-\mu p_1^2| \leq 4\mu-2$ for $\mu > 1$ (see [29]), we obtain

$$|a_3-2a_2^2| \leq \frac{8}{(2-\beta)^2}-\frac{2}{3-2\beta}.$$

Using this bound of $|a_3-2a_2^2|$ and the bounds of $|a_2|$, $|a_3|$ from (2.1) in (3.1), required bound of $|T_{3,1}(f)|$ follows. Sharpness of the bound of $|T_{3,1}(f)|$ follows from the function $\tilde{f}_1$.

Remark 3.1. The bounds of $|T_{2,n}(f)|$ and $|T_{3,1}(f)|$ for the class $\mathcal{R}$ follow from Theorem 3.1, when $\beta = 0$ [2, Theorem 2.12].

Theorem 3.2. *If $f \in \mathcal{A}_\beta$ is of the form (1.1), then*

$$T_{3,1}(f) \leq \begin{cases} \dfrac{4\beta^4-28\beta^3+37\beta^2-4\beta-4}{(3-2\beta)^2(2-\beta)^2}; & \frac{10-\sqrt{10}}{9} \leq \beta \leq 1, \\ 1; & 0 \leq \beta \leq \frac{10-\sqrt{10}}{9}. \end{cases} \tag{3.2}$$

The bounds are sharp.

Proof. For $f(z) = z+\sum_{n=2}^{\infty} a_n z^n \in \mathcal{A}_\beta$, Theorem 2.1 yields

$$|a_2| \leq \frac{2}{2-\beta} \text{ and } |a_3| \leq \frac{2}{3-2\beta}.$$

Hence $|a_2| \in [0,2]$ and $|a_3| \in [0,2]$ for $\beta \in [0,1]$. From (1.4), we have

$$\begin{aligned} T_{3,1}(f) &= 1+2\,\mathrm{Re}(a_2^2\bar{a}_3)-2|a_2|^2-|a_3|^2 \\ &\leq 1+2|a_2|^2|a_3|-2|a_2|^2-|a_3|^2 =: g(|a_3|), \end{aligned}$$

where $g(x) = 1+2|a_2|^2x-2|a_2|^2-x^2$ with $x = |a_3| \in [0,2]$. Since $g'(x) = 0$ at $x_0 := |a_2|^2$ and $g''(x_0) < 0$, therefore $g(x)$ attains its maximum value at $x = x_0$, whenever $|a_2|^2$ belongs to the range of x, that means $|a_2|^2 \leq 2$. Thus

$$\begin{aligned} T_{3,1}(f) \leq g(|a_2|^2) &= (|a_2|^2-1)^2 \\ &\leq 1 \quad \text{when} \quad |a_2|^2 \leq 2, \end{aligned}$$

$$= 1 \quad \text{when} \quad 0 \le \beta \le 2 - \sqrt{2}.$$

Now, the other case, when $|a_2|^2$ does not lie in the range of x, that is $|a_2|^2 > 2$ or $2 - \sqrt{2} \le \beta \le 1$, then

$$\begin{aligned} T_{3,1}(f) &\le \max g(x) = g\left(\frac{2}{3-2\beta}\right) \\ &= 1 - 2a_2^2 - \frac{4}{(3-2\beta)^2} + \frac{4a_2^2}{3-2\beta} \\ &\le \frac{4\beta^4 - 28\beta^3 + 37\beta^2 - 4\beta - 4}{(3-2\beta)^2(2-\beta)^2}. \end{aligned}$$

Using all these above arguments, we obtain

$$T_{3,1}(f) \le \begin{cases} 1, & 0 \le \beta \le \beta_0; \\ \frac{4\beta^4 - 28\beta^3 + 37\beta^2 - 4\beta - 4}{(3-2\beta)^2(2-\beta)^2}, & \beta_0 \le \beta \le 1, \end{cases}$$

where $\beta_0 = (10 - \sqrt{10})/9$ is the root of the equation $9\beta^2 - 20\beta + 10 = 0$.

The sharpness of the bound follows from $f(z) = z$ when $0 \le \beta \le (10 - \sqrt{10})/9$. However, for $(10 - \sqrt{10})/9 \le \beta \le 1$, equality in (3.2) holds for the function $\tilde{f}$ given in (2.3).

Remark 3.2. For $\beta = 0$ in Theorem 3.2, we obtain $T_{3,1}(f) \le 1$ for $f \in \mathcal{R}$ [20, Example 2.4].

Theorem 3.3. *If $f \in \mathcal{A}_\beta$ is of the form (1.1), then*

$$T_{3,1}(f) \ge 1 - \frac{4\beta - 9}{\beta^4 - 4\beta^3 + 2\beta^2 + 8\beta - 8}.$$

The bound is sharp.

Proof. Let $f \in \mathcal{A}_\beta$, then from (2.2), we have

$$a_2 = \frac{p_1}{2-\beta}, \quad a_3 = \frac{p_2}{3-2\beta}.$$

Now, by replacing p_2 in terms of p_1 using (2.6), we get

$$\begin{aligned} 2\operatorname{Re}(a_2^2 \bar{a}_3) &= \frac{p_1^4 + p_1^2(4 - p_1^2)\operatorname{Re}\zeta}{(3-2\beta)(2-\beta)^2}, \quad -|a_2|^2 = \frac{|p_1|^2}{(2-\beta)^2}, \\ -|a_3|^2 &= -\frac{p_1^4 + (4-p_1^2)^2|\zeta|^2 + 2p_1^2(4-p_1^2)\operatorname{Re}\zeta}{4(3-2\beta)^2}. \end{aligned}$$

A simple computation yields that

$$\begin{aligned} T_{3,1}(f) = 1 + \frac{1}{4(3-2\beta)^2(2-\beta)^2}\Big(& p_1^4(8 - 4\beta - \beta^2) - 8p_1^2(3-2\beta)^2 \\ & - (4-p_1^2)^2(2-\beta)^2|\zeta|^2 + 2p_1^2(4-p_1^2)(2-\beta^2)\operatorname{Re}\zeta\Big) \\ =: g(p_1, \zeta, \operatorname{Re}(\zeta)). \end{aligned}$$

Since the classes $\mathcal{A}_\beta$ and $\mathcal{P}$ are rotationally invariant, we can take $p = p_1 \in [0, 2]$. Using $\operatorname{Re}(\zeta) \ge -|\zeta|$ with notation $|\zeta| = y$, we have $g(p_1, |\zeta|, \operatorname{Re}\zeta) \ge g_1(p, y)$, where

$$\begin{aligned} g_1(p, y) = 1 + \frac{1}{4(3-2\beta)^2(2-\beta)^2}\Big(& p^4(8 - 4\beta - \beta^2) - 8p^2(3-2\beta)^2 \\ & - (4-p^2)^2(2-\beta)^2y^2 - 2p^2(4-p^2)(2-\beta^2)y\Big). \end{aligned}$$

Also, note that

$$\frac{\partial g_1(p,y)}{\partial y} = -\frac{2(4-p^2)^2y(2-\beta)^2 + 2p^2(4-p^2)(2-\beta^2)}{4(3-2\beta)^2(2-\beta)^2} < 0$$

for all $p \in [0,2]$ and $\beta \in [0,1]$. Hence $g_1(p,y)$ is a decreasing function of y with $g_1(p,y) \leq g_1(p,1) =: g_2(p)$. Minimum of $g_2(p)$ is the lower bound of $\det T_{3,1}(f)$. The equation $g_2'(p) = 0$ gives the following critical points

$$p^{(1)} = 0, \quad p^{(2)} = \pm\sqrt{\frac{(2\beta^2 - 8\beta + 7)}{(2-\beta^2)}}.$$

Using the basic calculus rule, it can be easily observed that the function $g_2(p)$ attains its minimum value at $p^{(2)}$ as $g''(p^{(2)}) > 0$ for all $\beta \in [0,1]$. Thus

$$\det T_{3,1}(f) \geq g_2(p^{(2)}) = 1 - (4\beta - 9)/(\beta^4 - 4\beta^3 + 2\beta^2 + 8\beta - 8).$$

To show the sharpness consider the function $\tilde{f}_3 \in \mathcal{A}$ given by

$$\beta\frac{\tilde{f}_3(z)}{z} + (1-\beta)\tilde{f}_3'(z) = \frac{1-z^2}{1 - z\sqrt{(2\beta^2 - 8\beta + 7)/(2-\beta^2)} + z^2}.$$

For $\tilde{f}_3(z) = z + \sum_{n=2}^{\infty} a_n z^n$, we have

$$a_2 = \frac{1}{2-\beta}\sqrt{\frac{2\beta^2 - 8\beta + 7}{2-\beta^2}}, \quad a_3 = \frac{1-2\beta}{2-\beta^2}$$

and $T_{3,1}(\tilde{f}_3) = 1 - (4\beta - 9)/(\beta^4 - 4\beta^3 + 2\beta^2 + 8\beta - 8)$.

Remark 3.3. For $\beta = 0$ in Theorem 3.2, we obtain $\det T_{3,1}(f) \geq -1/8$ for $f \in \mathcal{R}$ [20, Example 2.4].

4 Coefficient Difference

Robertson [34] proved that $3|a_{n+1} - a_n| \leq (2n+1)|a_2 - 1|$ for the class of convex functions. Recently, Li and Sugawa [25] obtained the bound of $|a_{n+1} - a_n|$ for particular choices of n for the class of convex function with fixed second coefficient. In this section, we find the the bound of $|a_{n+1}^N - a_n^N|$ $(N \in \mathbb{N})$ depending on the second coefficient for $f \in \mathcal{A}_\beta$. In fact, it is more convenient to express our result in terms of $p_1 = p$, applying the correspondence

$$(2-\beta)a_2 = p_1 = p.$$

To make the results more legible, we define the class $\mathcal{A}_\beta(p)$, $p \in [-2,2]$ as follows

$$\mathcal{A}_\beta(p) = \{f \in \mathcal{A}_\beta : f''(0) = p\}.$$

Clearly,

$$\bigcup_{-2\leq p\leq 2} \mathcal{A}_\beta(p) \subset \mathcal{A}_\beta \quad \text{and} \quad \bigcup_{-2\leq p\leq 2} \mathcal{A}_\beta(p) \neq \mathcal{A}_\beta.$$

The following lemmas are used to establish our main results.

Lemma 4.1. *[9] If $p(z) = 1 + \sum_{n=1}^{\infty} p_n z^n \in \mathcal{P}$, then the following estimate holds:*

$$|p_{n+1}^N - p_n^N| \leq 2^N\sqrt{2 - 2^{1-N}\operatorname{Re}(p_1^N)} \quad (N \in \mathbb{N}).$$

Equality holds for the function $(1 + e^{i\alpha}z)/(1 - e^{i\alpha}z)$, where $\alpha = \cos^{-1}(b/2)$ and $\operatorname{Re} p_1 = 2b$.

Lemma 4.2. *[23] Fix $\zeta \in \bar{\mathbb{D}}$. If $p(z) = 1 + \sum_{n=1}^{\infty} p_n z^n \in \mathcal{P}$, then*

$$|\xi p_{n+1} - p_n| \leq \frac{2(1-|\xi|^n)\left(1 + |\xi|^2 - \operatorname{Re}(\xi p_1)\right)}{1-|\xi|} + |2 - \xi p_1||\xi|^n \ \text{ for } \ |\xi| < 1.$$

The bounds are sharp for $p(z) = (1+z)/(1-z)$.

According to Komatu [19], if $p(z) = 1 + \sum_{n=1}^{\infty} p_n z^n$ and $q(z) = 1 + \sum_{n=1}^{\infty} q_n z^n$ both are the members of $\mathcal{P}$, then the weighted Hadamard product, $f * g$, also belongs to $\mathcal{P}$, where

$$f * g = 1 + \sum_{n=1}^{\infty} \frac{p_n q_n}{2} z^n.$$

Let us define $F_j(z) = F_{j-1} * p(z)$ for $j \in \mathbb{N}$ with $F_0(z) = p(z)$, then using the above result, we have $F_j \in \mathcal{P}$. Particulary, for $N \in \mathbb{N}$, the function

$$F_{N-1}(z) = 1 + \sum_{n=1}^{\infty} \frac{p_n^N}{2^{N-1}} z^n \in \mathcal{P}.$$

Replacing $p(z)$ in Lemma 4.2 by F_{N-1}, the result is as follows:

Lemma 4.3. *Fix $\xi \in \bar{\mathbb{D}}$ and $N \in \mathbb{N}$. If $p(z) = 1 + \sum_{n=1}^{\infty} p_n z^n$, then*

$$|\xi p_{n+1}^N - p_n^N| \leq \frac{2(1-|\xi|^n)\left(2^{N-1} + 2^{N-1}|\xi|^2 - \mathrm{Re}(\xi p_1^N)\right)}{1-|\xi|} + |2^N - \xi p_1^N| \cdot |\xi|^n \qquad \text{for } |\xi| < 1,$$

Equality holds for the function $(1+z)/(1-z)$.

Theorem 4.1. *If $f \in \mathcal{A}_\beta(p)$, then the following inequalities hold:*

$$|a_{n+1}^N - a_n^N| \leq \begin{cases} \dfrac{2(\sigma^n - \mu^n)(2^{N-1}\sigma^2 + 2^{N-1}\mu^2 - \sigma\mu p^N)}{(\sigma-\mu)\sigma\mu^{n+1}} + \dfrac{\sigma^n |2^N \mu - \sigma p^N|}{\sigma\mu^{n+1}}; & \beta \in [0,1), \\ \dfrac{2^N\sqrt{2 - 2^{1-N}p^N}}{\sigma}; & \beta = 1, \end{cases} \tag{4.1}$$

where $\sigma = (n-(n-1)\beta)^N$ and $\mu = (n+1-n\beta)^N$. Bounds for $\beta \in [0,1)$ is sharp for $p = 2$ whereas for $\beta = 1$, bound is sharp for odd N and $p = -2$.

Proof. For $f \in \mathcal{A}_\beta(p)$, from (2.2), we have

$$(n-(n-1)\beta)^N |a_{n+1}^N - a_n^N| = \left| \left(\frac{n-(n-1)\beta}{(n+1)-n\beta} \right)^N p_n^N - p_{n-1}^N \right|.$$

From Lemma 4.3 with $((n-(n-1)\beta)/((n+1)-n\beta))^N =: \xi$, bound in (4.1) for $\beta \in [0,1)$ follows since $\xi \in (0,1)$ whenever $\beta \in (0,1)$. For $\beta = 1$ we have $\xi = 1$. Bounds for $\beta = 1$ are obtained using Lemma 4.1 .

To show the sharpness for $\beta \in [0,1)$, consider the function $\tilde{f}(z)$ given in (2.3). As for $\tilde{f}$, we have

$$|a_{n+1} - a_n| = \frac{2^N}{(n-(n-1)\beta)^N} \left| \frac{(n-(n-1)\beta)^N}{(n+1-n\beta)^N} - 1 \right|,$$

which is same as in (4.1) for $p = 2$. In case of $\beta = 1$, for the function $\tilde{f}(-z)$, we have

$$|a_{n+1} - a_n| = \frac{2^{N+1}}{(n-(n-1)\beta)^N},$$

which coincides with the bounds in (4.1) for odd N and $p = -2$.

For $N = 1$, Theorem 4.1 yields the following bounds:

Corollary 4.2. *If $f \in \mathcal{A}_\beta(p)$ is of the form (1.1), then*

$$|a_{n+1}-a_n| \leq \begin{cases} \dfrac{2(\sigma^n-\mu^n)(\sigma^2+\mu^2-\sigma\mu p)}{(\sigma-\mu)\sigma\mu^{n+1}} + \dfrac{\sigma^n|2\mu-\sigma p|}{\sigma\mu^{n+1}}; & \beta \in [0,1), \\ \dfrac{2\sqrt{2-p}}{\sigma}; & \beta = 1. \end{cases}$$

The class $\mathcal{A}_\beta$ reduces to the class $\mathcal{R}$ for $\beta = 0$. Let us take corresponding class $\mathcal{R}(p) = \{f \in \mathcal{R} : f''(0) = p\}$. Theorem 4.1 gives the following result for the class $\mathcal{R}(p)$ when $\beta = 0$.

Corollary 4.3. *If $f \in \mathcal{R}(p)$ is of the form (1.1), then the following sharp bounds hold:*

$$|a_{n+1}^N - a_n^N| \leq \frac{2(\sigma^n-\mu^n)(2^{N-1}\sigma^2+2^{N-1}\mu^2-\sigma\mu p^N)}{(\sigma-\mu)\sigma\mu^{n+1}} + \frac{\sigma^n|2^N\mu-\sigma p^N|}{\sigma\mu^{n+1}}.$$

5 Growth Theorem and Bohr Phenomenon

Theorem 5.1. *If $f \in \mathcal{A}_\beta$ is of the form (1.1), then for $|z| \leq r$, the following hold:*

(i) $-\dfrac{\tilde{f}(-r)}{r} \leq \mathrm{Re}\left(\dfrac{f(z)}{z}\right) \leq \dfrac{\tilde{f}(r)}{r}$,

(ii) $-\tilde{f}(-r) \leq |f(z)| \leq \tilde{f}(r)$,

where $\tilde{f}(z)$ is given by (2.3). All these estimations are sharp.

Proof. (i) Let $f \in \mathcal{A}_\beta$. Consider $p(z) = f(z)/z$, then we have

$$\mathrm{Re}(p(z) + (1-\beta)zp'(z)) > 0.$$

It can be viewed as $p(z) + (1-\beta)zp'(z) \prec (1+z)/(1-z)$. Further, by Hallenbeck and Rusheweyeh [30, Theorem 3.1b], it follows that

$$p(z) \prec q(z) \prec \frac{1+z}{1-z},$$

where $q(z)$ is convex and best dominant, given by

$$\begin{aligned} q(z) &= \frac{1}{(1-\beta)z^{(\frac{1}{1-\beta})}} \int_0^z \left(\frac{1+t}{1-t}\right) t^{(\frac{1}{1-\beta}-1)} dt \\ &= \frac{\tilde{f}(z)}{z}, \end{aligned}$$

where $\tilde{f}(z)$ is defined in (2.3). Since $q(z)$ is convex and all coefficients are real for $\beta \in [0,1]$, therefore image domain of $\mathbb{D}$ under the function $q(z)$ is symmetric with respect to real axis and

$$q(-r) \leq \mathrm{Re}(q(z)) \leq q(r), \quad |z| = r < 1.$$

As $p(z) = f(z)/z \prec q(z)$, so required bound of $\mathrm{Re}(f(z)/z)$ follows. This completes the first part. Sharpness of the bounds follow as $q(z)$ is the best dominant. (ii) From [7, Lemma 4.10], $f \in \mathcal{A}_\beta$ if and only if

$$f(z) = z\int_0^1 p(t^{1-\beta}z)dt, \tag{5.1}$$

where $p \in \mathcal{P}$. Using the well known bound $|p(z)| \leq (1+r)/(1-r)$ of Carathéodory functions, we have

$$|f(z)| \leq r\int_0^1 \frac{1+rt^{1-\beta}}{1-rt^{1-\beta}}dt = \tilde{f}(r),$$

Now, we proceed for the lower bound of $|f(z)|$. After solving the integration in (5.1) for $p(z) = (1+z)/(1-z)$, we get

$$f(z) = z(-1 + 2H(z)),$$

where

$$H(z) = {}_2F_1\left[1, \frac{1}{1-\beta}, \frac{2-\beta}{1-\beta}, z\right].$$

Thus for $z = re^{i\theta}$,

$$|f(z)| = |z(-1 + 2H(z))| \geq \min_{\theta\in[0,2\pi]} g(\theta) \tag{5.2}$$

where

$$g(\theta) = \sqrt{\mathrm{Re}(re^{i\theta}(-1 + 2H(re^{i\theta})))^2 + \mathrm{Im}(re^{i\theta}(-1 + 2H(re^{i\theta})))^2},$$

Since for different choices of β in $[0, 1)$, $H(z)$ reduces to different functions. For instance, when $\beta = 0$, it becomes $-2\log(1-z)/z$ and for $\beta = 1/2$, it reduces to $-4(z + \log(1-z))/z^2$. By a simple calculation, we find that the function $g(\theta)$ is decreasing from $[0, \pi]$ and increasing from $[\pi, 2\pi]$ for $r \in (0, 1)$ and $\beta \in [0, 1)$. Hence $g(\theta)$ attains its minimum value at $\theta = \pi$. Thus from (5.2), we get

$$\begin{aligned} |f(z)| &\geq |-r(-1 + 2H(-r))| \\ &= r(-1 + 2H(-r)) = -r\tilde{f}(-r), \end{aligned}$$

which completes the proof. Bounds are sharp for the function $\tilde{f}(z)$.

Theorem 5.2. *If $f \in \mathcal{A}_\beta$ is of the form (1.1), then for $m \in \mathbb{N}$*

$$|\omega(z^m)| + \sum_{n=2}^{\infty} |a_n z^n| \leq d(0, \partial f(\mathbb{D}))$$

in $|z| \leq r^*$, where r^* is the smallest positive root of

$$r^m + \tilde{f}(r) - r + \tilde{f}(-1) = 0. \tag{5.3}$$

The radius r^* is sharp.

Proof. Let $f \in \mathcal{A}_\beta$, then by Theorem 5.1, the Euclidean distance between $f(0) = 0$ and the boundary of $f(\mathbb{D})$ satisfies

$$d(0, \partial f(\mathbb{D})) \geq \lim_{r\to 1} |f(z)| = -\tilde{f}(-1).$$

Let $|z| \leq r$. Now using (2.1) with the above inequality, we have

$$\begin{aligned} |\omega(z^m)| + \sum_{n=2}^{\infty} |a_n z^n| &\leq r^m + \sum_{n=2}^{\infty} \left(\frac{2}{n - \beta(n-1)}\right) r^n, \\ &= r^m + \tilde{f}(r) - r \\ &\leq -\tilde{f}(-1) \leq d(0, \partial f(\mathbb{D})). \end{aligned}$$

which is true in $|z| = r \leq r^*$, where r^* is the root of $H(r) = r(r^{m-1} - 1) + \tilde{f}(r) + \tilde{f}(-1)$. Note that, $H(0) = \tilde{f}(-1) < 0$ and $H(1) = \tilde{f}(1) + \tilde{f}(-1) > 0$ for all $\beta \in [0, 1]$, therefore by the Intermediate value property for continuous functions there must exist a $r^* \in (0, 1)$ such that $H(r^*) = 0$.

Sharpness holds for the functions $\tilde{f}(z)$ and $\omega(z) = z$. Since at $z = r^*$,

$$\begin{aligned} |\omega(z^m)| + \sum_{n=2}^{\infty} |a_n z^n| &= (r^*)^m + \sum_{n=2}^{\infty} \frac{2}{n - (n-1)\beta} (r^*)^n \\ &= (r^*)^m + \tilde{f}(r^*) - r^* = -\tilde{f}(-1). \end{aligned}$$

Hence the radius is sharp.

For $w(z) = z$ and $m = 1$, Theorem 5.2 gives the following Bohr-radius for the class $\mathcal{A}_\beta$.

Corollary 5.3. *If $f \in \mathcal{A}_\beta$, then $|z|+\sum_{n=2}^{\infty}|a_n z^n| \leq d(0, \partial f(\mathbb{D}))$ in $|z| \leq r_b$, where r_b is root of $\tilde{f}(r)+\tilde{f}(-1) = 0$. The radius r_b is sharp.*

For various values of $\beta \in [0, 1]$, the root r_b is shown in Figure 1 and Table 1.

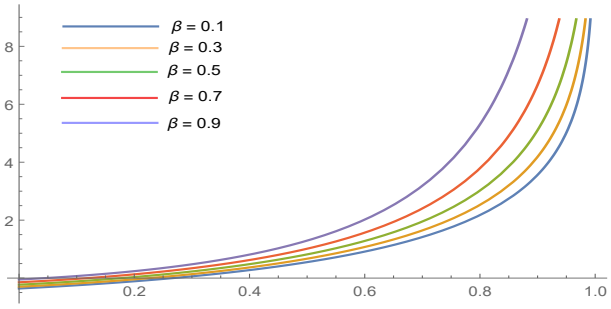


Figure 1

Table 1: Radius r^* for various choices of β

β	0.1	0.2	0.3	0.5	0.7	0.8	0.9
r_b	0.267139	0.24766	0.22655	0.178366	0.119726	0.085113	0.0457777

Theorem 5.4. *If $f \in \mathcal{A}_\beta$, then*

$$|f(z^m)| + \sum_{k=N}^{\infty}|a_k z^k| \leq d(0, \partial f(\mathbb{D})) \tag{5.4}$$

hold for $|z| = r \leq r_N$, where r_N is the root of the equation

$$\tilde{f}(r^m) + \tilde{f}(r) - \hat{f}(r) + \tilde{f}(-1) = 0,$$

with

$$\hat{f}(r) = \begin{cases} 0 & N = 1, \\ r & N = 2, \\ r + \sum_{n=2}^{N-1} \frac{2}{n-(n-1)\beta} r^n & N \geq 3. \end{cases}$$

The radius is sharp.

Proof. Suppose $f \in \mathcal{A}_\beta$, then from (1.2) and Theorem 5.1, we have

$$\begin{aligned} |f(z^m)| + \sum_{k=N}^{\infty}|a_k z^k| &\leq \tilde{f}(r^m) + \sum_{n=N}^{\infty} \frac{2}{n-(n-1)\beta} r^n \\ &= \tilde{f}(r^m) - \hat{f}(r) + \tilde{f}(r) \\ &\leq -\tilde{f}(-1) \\ &\leq d(0, \partial f(\mathbb{D})) \end{aligned}$$

holds in $|z| = r_N$, where r_N is the root of

$$G(r) := \tilde{f}(r^m) - \hat{f}(r) + \tilde{f}(r) + \tilde{f}(-1) = 0.$$

Since $G(0) = \tilde{f}(-1) < 0$ and $G(1) = (\tilde{f}(1) - \hat{f}(1)) + (\tilde{f}(1) + \tilde{f}(-1)) > 0$, therefore there exist a $r_N \in (0,1)$ such that (5.4) holds. Note that, for the function $\tilde{f}(z)$ at $|z| = r_N$,

$$\begin{aligned} |f(z^m)| + \sum_{k=N}^{\infty} |a_k z^k| &= \tilde{f}((r_N)^m) + \sum_{n=N}^{\infty} \frac{2}{n-(n-1)\beta} (r_N)^n \\ &= -\tilde{f}(-1), \end{aligned}$$

which proves the sharpness of radius.

6 Declarations

Funding

The work of the Surya Giri is supported by University Grant Commission, New-Delhi, India under UGC-Ref. No. 1112/(CSIR-UGC NET JUNE 2019).

Conflict of interest

The authors declare that they have no conflict of interest.

Author Contribution

Each author contributed equally to the research and preparation of manuscript.

Data Availability

Not Applicable.

References

[1] V. Allu, A. Lecko and D. K. Thomas, Hankel, Toeplitz, and Hermitian-Toeplitz determinants for certain close-to-convex functions, Mediterr. J. Math. **19** (2022), no. 1, Paper No. 22, 17 pp.

[2] M. F. Ali, D. K. Thomas and A. Vasudevarao, Toeplitz determinants whose elements are the coefficients of analytic and univalent functions, Bull. Aust. Math. Soc. **97** (2018), no. 2, 253–264.

[3] E. Berkson and H. Porta, Semigroups of analytic functions and composition operators, Michigan Math. J. **25** (1978), no. 1, 101–115.

[4] C. Bénéteau, A. Dahlner and D. Khavinson, Remarks on the Bohr phenomenon, Comput. Methods Funct. Theory **4** (2004), no. 1, 1–19.

[5] H. Bohr, A Theorem Concerning Power Series, Proc. London Math. Soc. (2) **13** (1914), 1–5.

[6] F. Bracci, M. D. Contreras and S. Díaz-Madrigal, Pluripotential theory, semigroups and boundary behavior of infinitesimal generators in strongly convex domains, J. Eur. Math. Soc. (JEMS) **12** (2010), no. 1, 23–53.

[7] F. Bracci, M. D. Contreras, S. Díaz-Madrigal, M. Elin and D. Shoikhet, Filtrations of infinitesimal generators, Funct. Approx. Comment. Math. **59** (2018), no. 1, 99–115.

[8] F. Bracci, M. D. Contreras and S. Díaz-Madrigal, Evolution families and the Loewner equation I: the unit disc, J. Reine Angew. Math. **672** (2012), 1–37.

[9] J. E. Brown, Successive coefficients of functions with positive real part, Int. J. Math. Anal. (Ruse) **4** (2010), no. 49-52, 2491–2499.

[10] N. E. Cho, O. S. Kwon, A. Lecko and Y. J. Sim, Sharp estimates of generalized Zalcman functional of early coefficients for Ma-Minda type functions, Filomat **32** (2018), no. 18, 6267–6280.

[11] K. Cudna, O. S. Kwon, A. Lecko, Y. J. Sim and B. Śmiarowska, The second and third-order Hermitian Toeplitz determinants for starlike and convex functions of order α, Bol. Soc. Mat. Mex. (3) **26** (2020), no. 2, 361–375.

[12] P. L. Duren, *Univalent functions*, Grundlehren der mathematischen Wissenschaften, 259, Springer-Verlag, New York, 1983.

[13] M. Elin and D. Shoikhet, *Linearization models for complex dynamical systems*, Operator Theory: Advances and Applications, 208, Birkhäuser Verlag, Basel, 2010.

[14] M. Elin, D. Shoikhet and N. Tuneski, Radii problems for starlike functions and semigroup generators, Comput. Methods Funct. Theory **20** (2020), no. 2, 297–318.

[15] K. Gangania and S. S. Kumar, Bohr Radius for Some Classes of Harmonic Mappings, Iranian Journal of Science and Technology, Transactions A: Science (2022): 1-8.

[16] S. Giri and S. S. Kumar, Radius and Convolution problems of analytic functions involving Semigroup Generators, arXiv:2205.10777 (2022).

[17] K. Gangania and S. S. Kumar, Bohr-Rogosinski Phenomenon for $\mathcal{S}^*(\psi)$ and $\mathcal{C}(\psi)$, Mediterr. J. Math. **19** (2022), no. 4, Paper No. 161.

[18] I. R. Kayumov, D. M. Khammatova and S. Ponnusamy, Bohr-Rogosinski phenomenon for analytic functions and Cesáro operators, J. Math. Anal. Appl. **496** (2021), no. 2, Paper No. 124824, 17 pp

[19] Y. Komatu, On convolution of power series, Kodai Math. Sem. Rep. **10** (1958), 141–144.

[20] V. Kumar and N. E. Cho, Hermitian-Toeplitz determinants for functions with bounded turning, Turkish J. Math. **45** (2021), no. 6, 2678–2687.

[21] V. Kumar and S. Kumar, Bounds on Hermitian-Toeplitz and Hankel determinants for strongly starlike functions, Bol. Soc. Mat. Mex. (3) **27** (2021), no. 2, Paper No. 55, 16 pp.

[22] E. Landau and D. Gaier, *Darstellung und Begründung einiger neuerer Ergebnisse der Funktionentheorie*, third edition, Springer-Verlag, Berlin, 1986.

[23] A. Lecko, On coefficient inequalities in the Carathéodory class of functions, Annales Polonici Mathematici. Vol. 75. Instytut Matematyczny Polskiej Akademii Nauk, 2000.

[24] A. Lecko, Y. J. Sim and B. Śmiarowska, The fourth-order Hermitian Toeplitz determinant for convex functions, Anal. Math. Phys. **10** (2020), no. 3, Paper No. 39, 11 pp.

[25] M. Li and T. Sugawa, A note on successive coefficients of convex functions, Comput. Methods Funct. Theory **17** (2017), no. 2, 179–193.

[26] R. J. Libera and E. J. Złotkiewicz, Early coefficients of the inverse of a regular convex function, Proc. Amer. Math. Soc. **85** (1982), no. 2, 225–230.

[27] M.-S. Liu, Y.-M. Shang and J.-F. Xu, Bohr-type inequalities of analytic functions, J. Inequal. Appl. **2018**, Paper No. 345, 13 pp.

[28] W. Ma, Generalized Zalcman conjecture for starlike and typically real functions, J. Math. Anal. Appl. **234** (1999), no. 1, 328–339.

[29] W. C. Ma and D. Minda, A unified treatment of some special classes of univalent functions, in *Proceedings of the Conference on Complex Analysis (Tianjin, 1992)*, 157–169, Conf. Proc. Lecture Notes Anal., I, Int. Press, Cambridge, MA.

[30] S. S. Miller and P. T. Mocanu, *Differential subordinations*, Monographs and Textbooks in Pure and Applied Mathematics, 225, Marcel Dekker, Inc., New York, 2000.

[31] Y. A. Muhanna, R. M. Ali and S. Ponnusamy, On the Bohr inequality, in *Progress in approximation theory and applicable complex analysis*, 269–300, Springer Optim. Appl., 117, Springer, Cham.

[32] Y. A. Muhanna, Bohr's phenomenon in subordination and bounded harmonic classes, Complex Var. Elliptic Equ. **55** (2010), no. 11, 1071–1078.

[33] M. Obradović and N. Tuneski, Hermitian Toeplitz determinants for the class $\mathcal{S}$ of univalent functions, Armen. J. Math. **13** (2021), Paper No. 4, 10 pp.

[34] M. S. Robertson, Univalent functions starlike with respect to a boundary point, J. Math. Anal. Appl. **81** (1981), no. 2, 327–345.

[35] W. Rogosinski, Üover image barriers in power series and their sections, Math. Z. **17** (1923), no. 1, 260–276.

[36] S. Reich and D. Shoikhet, The Denjoy–Wolff theorem, Math. Encyclopaedia, Supplement 3, Kluwer Academic Publishers, (2002), 121–123.

[37] D. Shoikhet, *Semigroups in geometrical function theory*, Kluwer Academic Publishers, Dordrecht, 2001.

[38] D. K. Thomas, N. Tuneski and A. Vasudevarao, *Univalent functions*, De Gruyter Studies in Mathematics, 69, De Gruyter, Berlin, 2018.

[39] L. Wu, Q. Wang and B. Long, Some Bohr-type inequalities with one parameter for bounded analytic functions, Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat. RACSAM **116** (2022), no. 2, Paper No. 61, 13 pp.

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India

E-mail address: spkumar@dtu.ac.in

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India

E-mail address: suryagiri456@gmail.com

Indian Journal of Pure & Applied Physics
Vol. 60, October 2022, pp. 841-848
DOI:10.56042/ijpap.v60i10.64990

Compact Dual-Band Zig Zag Shaped Implantable Antenna for Biomedical Devices

Rajiv Kumar Nehra[a*] & N S Raghava[b]

[a]Department of Electronics and Communication Engineering, Bharati Vidyapeeth's College of Engineering, New Delhi 110 063, India

[b]Department of Electronics and Communication Engineering, Delhi Technological University, New Delhi 110 042, India

Received 29 July 2022; accepted 16 September 2022

This paper presents a highly compact volume 12 mm^3 Zig Zag-shaped implantable antenna for biomedical applications. The antenna performance is closely analyzed in terms of penetration depth, different human tissues, substrate materials, reflection coefficient, Impedance Bandwidth, directive gain, and SAR values. The three different models are analyzed for different resonating frequencies. It is found that the proposed antenna shows dual ISM bands (900 MHz to 970 MHz 2.36GHz to 2.49 GHz) with impedance bandwidth of 7.77 % and 5.5% respectively. The proposed implantable antenna is successfully fabricated and tested in skin mimicking gel (vitro model). The performance of the proposed antenna is enhanced by using a sorting pin and slots that provides dual ISM bands, an acceptable SAR of 856.3 W/KG, and a directive gain of -28.68 dBi. Parametric analysis of the proposed implantable antenna has been studied in detail. Simulated and fabricated results are compared and found almost similar

Keywords: ISM (Industrial Scientific and medical); SAR (Specific absorption rate); BMD (Biomedical devices); IMD (implantable medical devices); MDAN (medical body area network)

1 Introduction

In today's world, health is the most important concern for everyone. If any patient goes to a hospital, he or she mustwait for a long time in a long queue for medical check-ups, especially in countries having high population. Many serious diseases are diagnosed by doctors at a later stage. This may lead to getting serious damage to patients. So now a day's implantable medical devices are emerging as a new hope for a healthy lifestyle. Such devices are implanted inside the human body, and all biological information of the human body like sugar, blood pressure, glucose level, heartbeat, the oxygen level in the blood, *etc* will be retrieved. This technology is extremely beneficial to human beings for maintaining their health and anticipating their internal organs update in ad- vance[1]. In biomedical devices, an implantable antenna is a key component to establish wireless communication between BMDs and external receiver station. For diagnosis purpose, BMD is implanted inside the human body, so the size of BMD is always a great concern. Further, it is known that the major size of BMD is covered by an implantable antenna itself, which attracts worldwide researchers to work on antennasize minimization[2]. So, miniaturization of an implantable antenna for BMD is highly demanded. Various techniques have been studied by researchers so far. Miniaturization is achieved by lengthening the current path on the patch sur- face. Due to this reason, researchers generally use meandered antenna[3] to get a small size. This approach is further modified by using a sorting pin that makes the patch and ground surface sorted to each other, which results in a large current path on the patch and ground surface. Eventually, such antennas (PIFA) shift their resonating frequency to lower side[4]. Since the antenna is inside the human body so the internal environment of the human body is of absorbing nature and hence degrades antenna performance. To get best performance from implantable antennas, is quite challenging task for researchers in lossy environment of human body[5]. So, to evaluate performance of implantable antenna, SAR is also an important aspect. The amount of heat absorbed by the body per kg of mass is defined as SAR. A very high absorption of antenna radiation may cause a serious problem (temperature rise) in the human body because the higher temperature may change the relative permittivity of human tissue[5,6].

*Corresponding authors:
(Email: rajiv.nehra@bharatividyapeeth.edu; nsraghava@dtu.ac.in)

One more important challenge is to provide a wireless charging facility for the existing battery in BMDs. All electronic devices work on DC biasing energy. The proper functioning of BMDs can only be ensured with charged DC battery. It is quite inconvenient for patients to charge or replace the battery of BMDs by making surgery every time. So, compatible methods for wireless charging of the battery are very much required. To achieve wireless charging, the implantable antenna must be designed with dual bands (ISM bands). Here, the two bands are utilized in the manner like one band is used for data transmission during telemetry sessions and the other band utilizes radiation and convert them into DC voltage using rectifier, filter (integral component of BMDs) for wireless charging without disturbing the position of IMDs in the human body. In this paper, a highly compact Zig Zag- shaped implantable antenna has been designed, simulated, and analyzed for different parameters. An acceptable range of SAR is achieved, which is quite low as compared to IEEE standard safety guidelines[7,8] along with dual bands (ISM frequency bands), and the highly miniatured volume of the implantable antenna. The designing consideration steps, parametric study, performance analysis, and experiment results of a highly compact Zig Zag-shaped implantable antenna have been provided in detail. The measurement of a fabricated proposed implantable antenna with detailed discussion has been carried out.

2 Design and Fabrication of The Proposed antenna

A very compact implantable antenna is proposed that has an extremely small size as compared to recent reference papers[8-15]. All designed dimensions of antenna are presented in Table 1. An implantable antenna is designed in X mm x Y mm size. The proposed antenna is having the shape of two symmetric serpentine paths with one centre strip which is resembled with Zig Zag shape hence named a highly compact Zig-Zag shaped implantable antenna as shown in Fig. 1.

Table 1 — Design Parameters of Proposed Antenna.

Parameters	Value (mm)	Parameters	Value (mm)
A	0.5	E	0.4
B	1.5	F	0.45
C	1	r	0.75
D	0.25	s	2.75
X	6	p	2.25
Y	4	q	1
H	0.25		

2.1 Designing sequence of the proposed antenna

1. Initially, a size of X mm x Y mm is considered for patch design. Then the concept of meandered planer antenna is adapted to provide a longer current path. The symmetrical serpentine shape is designed on both sides of the centre strip with A mm width shown in Fig. 1(a). After the optimization process, a good impedance matching is achieved at position (r, s) of coax feed. This 1st designed antenna shows the resonating frequency of 1.48 GHz with a reflection coefficient of -23.65 dB. The impedance bandwidth of 12.94% from 1.39 GHz to 1.57 GHz has been achieved in Fig. 1(f).

2. To achieve the resonating frequency in the ISM band from 1st designed antenna, a sorting pin with radius 0.2 mm is introduced between the patch and ground surface via the substrate. Inclusion of sorting pin in 1st designed antenna is termed as 2nd designed antenna shown in Fig. 1(b). Complete optimization of sorting pin position on the patch surface is studied in detail. At last, sorting pin position (p, q) is finalized. At this specific position, 2nd designed antenna turned into a dual-band antenna shown in Fig. 1(f). This antenna exhibits resonating frequencies of 0.94 GHz, and 2.34 GHz with impedance bandwidth of 3.2% and 5.26% respectively. The reflection coefficient at both resonating frequencies is -10.2 and -21.44 dB respectively.

3. To take both frequency bands in the standard frequency range (ISM Bands), 3 slots are embedded on the left side of patch surface in 2nd designed antenna shown in Fig. 1(c). With help of 3 slots, ISM Bands are achieved with sufficient impedance bandwidth. Now the proposed implantable antenna exhibits resonating frequencies of 0.93 GHz and 2.43 GHz with impedance bandwidth from 0.90 GHz to 0.97 GHz (7.77%) and 2.36 GHz to 2.49 GHz (5.5%) along with -14.14 dB and -35.70 dB reflection coefficient respectively shown in Fig. 1(f). The width of each strip except D in given Fig. 1(c) is A mm. This antenna is designed on substrate Rogers 6010 RT/Duroid that has a relative permittivity of 10.2 and dielectric loss tangent 0.0023. The same material (Rogers 6010 RT/Duroid) is also placed over a patch surface which is called a superstrate. A high dielectric superstrate makes the antenna more efficient towards stabilizing the

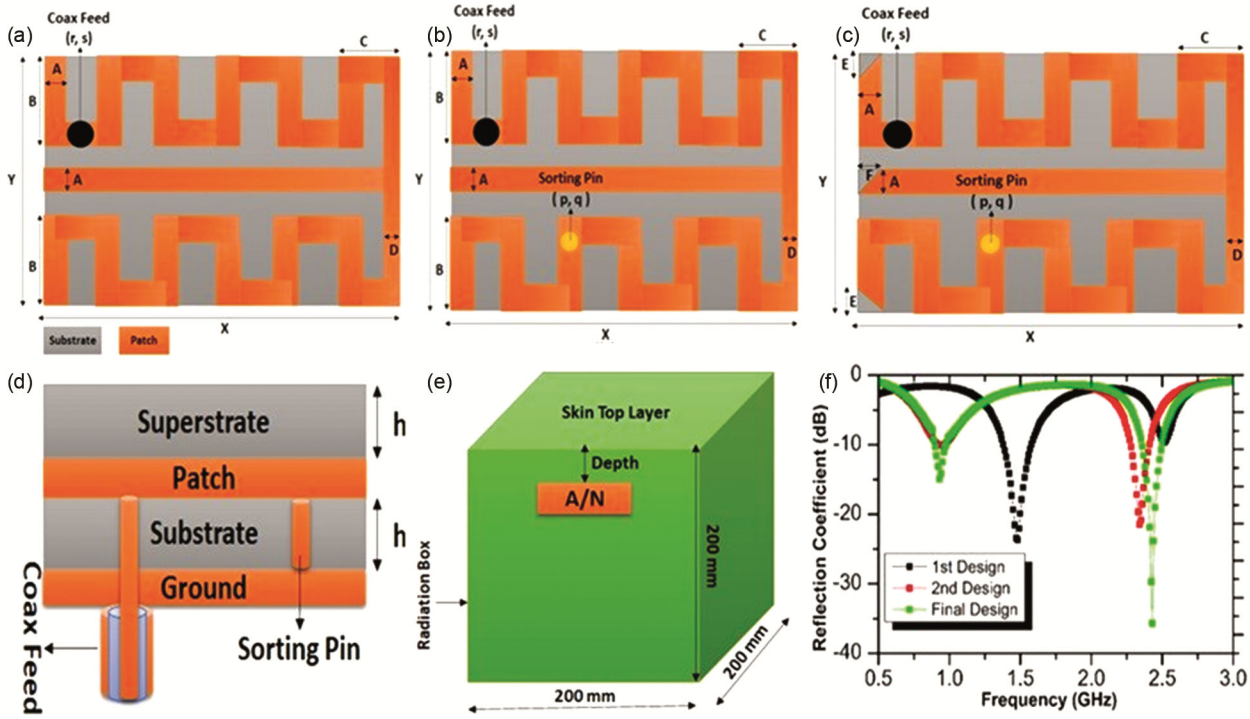


Fig. 1 — The designing sequence of Proposed antennas. (a) 1st designed antenna with feeding position (r, s). (b) 2nd designed antenna with sorting pin. (c) Design of Final proposed implantable antenna with slots. (d) Side view of the final proposed antenna with coaxial feed and sorting pin. (e) Placing of the final proposed antenna inside the skin box with a depth of 3.5 mm from the top surface. (f) Reflection coefficient of design sequence of antennas

fluctuation in relative permittivity and avoids antenna absorption from surrounding. Substrate and superstrate are selected in such a way to reduce reflection, better biocompatibility, and good impedance matching. Substrate and superstrate have a thickness of h mm each. So total volume of the proposed antenna is X mm x Y mm x 2*h mm (12 mm^3). The coaxial feed is used to excite the proposed antenna which has 50-ohm resistance. The position of the coaxial feed connector is at r, s position. At this point, the antenna exhibits good impedance matching. The final proposed antenna is placed at penetration depth of 3.5 mm inside human skin box in Fig. 1(e).

Ansys high-frequency structure simulator has been used for the analysis and designing of proposed research work. A skin box is created in the HFSS simulation tool by assigning a specific value of relative permittivity, bulk conductivity, and dimensions. The skin has unique properties like permittivity, bulk conductivity, and mass density based on the operating frequency. The proposed implantable antenna is designed in an ISM band (2.4 to 2.48 GHz), so the skin box is assigned with relative permittivity of 42.9 and bulk conductivity of 1.56 siemens per meter[17]. The volume of the skin box is 200 mm x 200 mm x 200 mm shown in Fig. 1(e).

3 Parametric Study of Proposed Antenna

The proposed implantable Antenna performance is exorbi- tantly investigated by varying the penetration depth in several tissues of the human body. Therefore, different tissues and depth of penetration of the proposed implantable antenna are deliberated in detail. Parametric study of various dielectric materials is also analyzed for the final antenna design.

3.1 Effect of penetration depth in skin model

Figure 2(a) represents the variation of reflection coefficient of the proposed implantable antenna at a different penetration depth from the top surface of the skin tissue model. During the investigation of penetration depth effect on proposed im- plantable antenna, antenna is placed at 6 different depths from the top surface of the skin tissue model. It is elucidated that the proposed antenna performs well in 2 ISM bands with a slight variation of resonating frequency. The penetration depth shows negligible effect on the performance of proposed antenna. At 3 mm penetration depth, the performance of

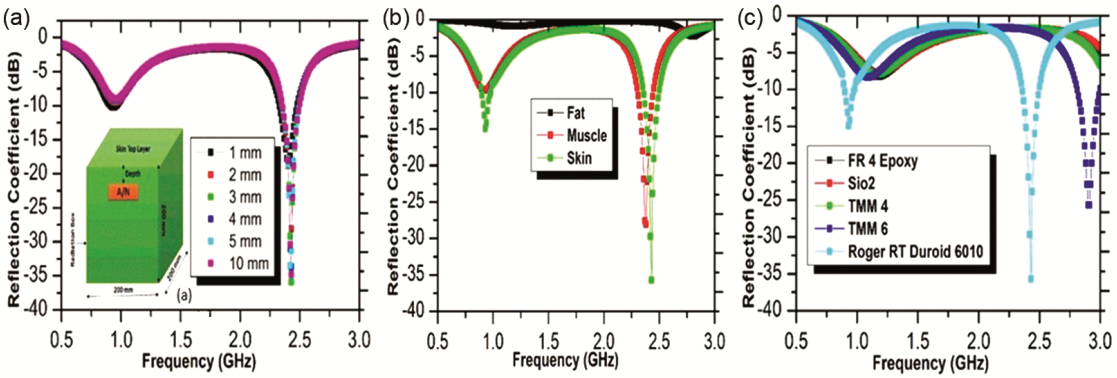


Fig. 2 — Reflection coefficient of proposed antenna at different penetration depths in skin tissue model. (b) Reflection coefficient of proposed antennain different human body tissues. (c) Reflection coefficient of the proposed antenna of different substrate materials.

proposed antenna upgrades in the lower ISM band (902 MHz to 928 MHz). It means the proposed antenna almost holds its performance (ISM Bands) with penetration depth.

3.2 Effect on antenna performance in various human body tissues

To validate the performance of the proposed antenna, it is placed in the vicinity of different human body tissues like Skin, Fat, Muscle, *etc*. The performance of the proposed antenna is evaluated in terms of reflection coefficient and is shown in Fig. 2(b). The reflection coefficient curves indicate that the proposed antenna performs excellent results in the ISM band when kept in the skin model, while in the fat and muscle layer, the antenna has lost a lower ISM band (S11-10 dB).In the skin layer, it radiates energy in dual bands with good impedance matching. It implies that from given human tissues, the proposed antenna is extremely suitable for skin tissues at 3.5 mm deep inside from the top surface.

3.3 Effect of substrate materials on antenna performance

Substrate material provides mechanical strength and robustness to any microwave antenna. Depending upon different dielectric properties, five substrates have been investigated for performance evaluation of the proposed antenna. The reflection coefficient performance is shown in Fig. 2(c). It is observed that the substrate material Rogers 6010 RT/Duroid exhibits outstanding performance in dual ISM bands. Furthermore, other substrates do not show good response in desired frequency bands. So, the selection of Rogers 6010 RT/Duroid substrate material is an excellent choice for the proposed implantable antenna.

3 Measurement and Results Discussion

The implantable antennas are generally tested in two different methodologies. One is vivo model in which human body phantom is carried out and implantable antenna is placed at specific position inside it. Other method (vitro model) is to develop skin mimicking gel solution which has same properties of human skin at specific operating frequency. In this section, to test the proposed antenna's performance, vitro model is adopted. Skin mimicking gel is depicted in Fig. 3(a). This liquid gel has properties equivalent to human skin of dielectric constant er=42.9, Bulk Conductivity of 1.56 Siemen/meter in the ISM band[16]. Specific quantity of sucrose 53%, deionised water 47%, and carbomer 0.5 gram in 40 mL solution, is mixed well and put that solution at 80 °C for one hour[17]. Then solution is left to cool down for some time shown in Fig. 3(a). The proposed antenna Measurement set up is shown in Fig. 3(c). The proposed implantable antenna is immersed with a depth of 3.5 mm in skin mimicking gel as shown in Fig. 3(c). KEYSIGHT Field Fox RF Analyser N9914A 6.5 GHz network analyzer and 50-ohm SMA connector are used to measure antenna parameters. The proposed antenna is fabricated on printed circuit board shown in Fig. 3(b). To install sorting pin, cautious efforts are taken place. A drilling machine with radius 0.2 mm is used to make sharp hole in PCB. Copper wire is inserted as sorting pin between patch and ground surface of proposed antenna. To develop the slots on patch surface, etching technique is used to remove specific conducting area using copper chloride solution. Then fabricated antenna is tested successfully in the microwave lab. Measured results have good agreement with simulated results shown in Fig. 3(e).

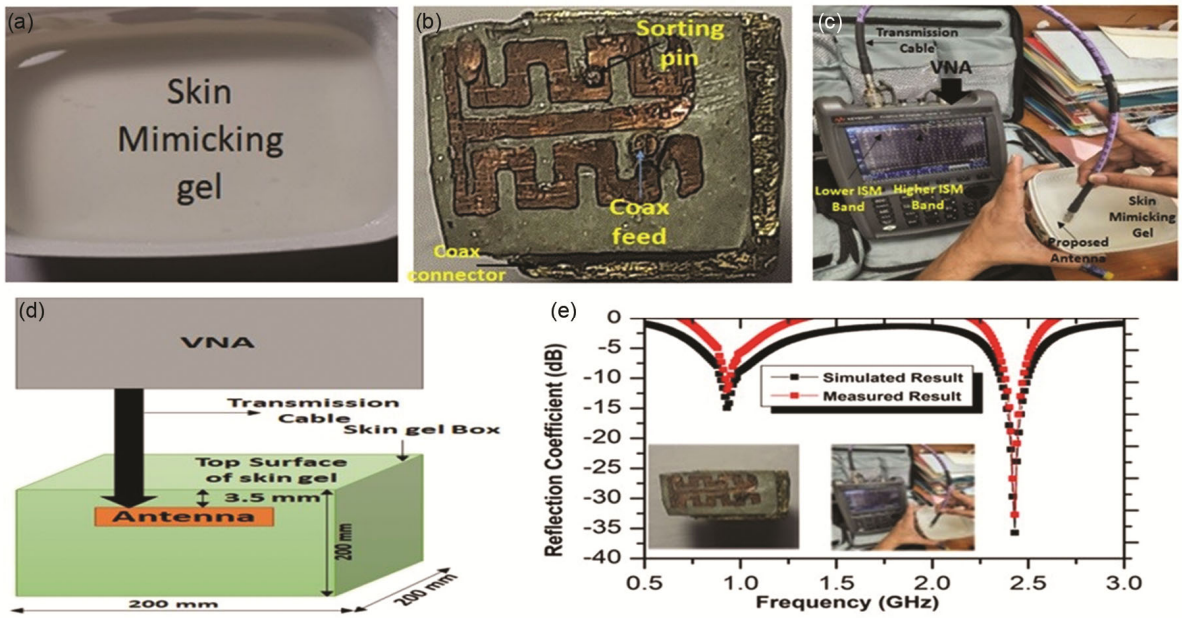


Fig. 3 — (a) Skin mimicking gel (vitro model) at 2.43 GHz. (b) fabrication of proposed implantable antenna. (c) Measurement sight of the fabricated antennain the microwave lab. (d) Prototype model of antenna result measurement. (e) Reflection coefficient of simulated and fabricated antenna. (e) .

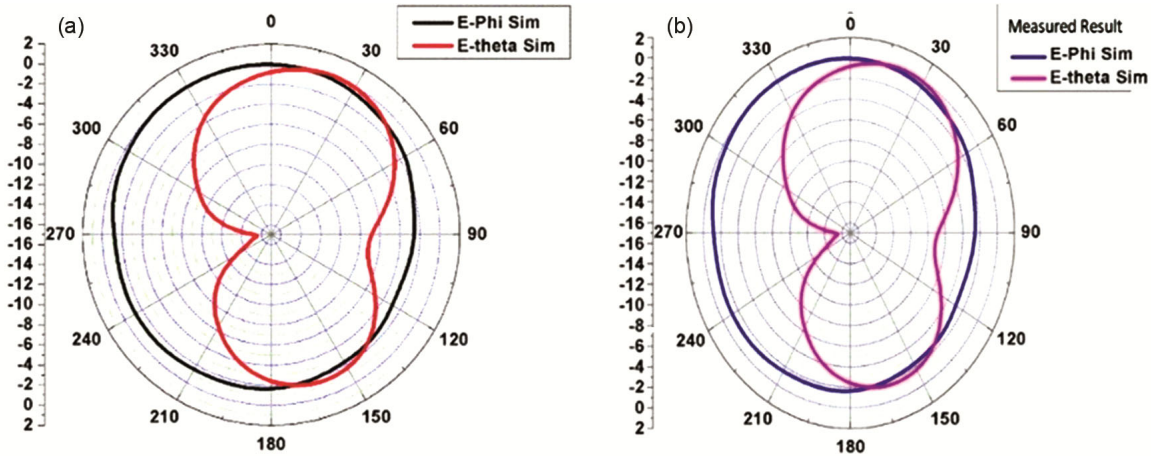


Fig. 4 — (a) Simulation of the Radiation pattern of the proposed antenna at 2.43 GHz. (b) Measurement of the radiation pattern of fabricated antenna at 2.43 GHz

The fabricated antenna is tested and measured reflection coefficient (dB) in skin gel with dual-band performance shown in Fig. 3(c). The fabricated antenna shows dual bands with slight variation in reflection coefficient. It has impedance bandwidth from 0.92 to 0.94 GHz and 2.38 to 2.47 GHz in lower and higher ISM bands respectively. However, the slight variation of the simulated and fabricated antenna in impedance bandwidth occurs due to limitation of fabrication technique, soldering technique, presence of air gap, imperfection of sorting pin, and glue between patch surface and superstrate. Prototype of measurement set up of implantable antenna is also presented in Fig. 3(d).

From Fig. 4, it is well observed that the simulated and fabricated antenna exhibits almost similar radiation pattern for gain performance. Moreover, with introduction of sorting and slots, proposed antenna design has improved gain value from - 30.30 dB (1st design antenna) to -28.68 dB. The negative value of gain shows that human body tissues absorb radiation from antenna section in BMDs. This absorption of antenna radiation leads to get negative value of gain. But this gain value is sufficient to

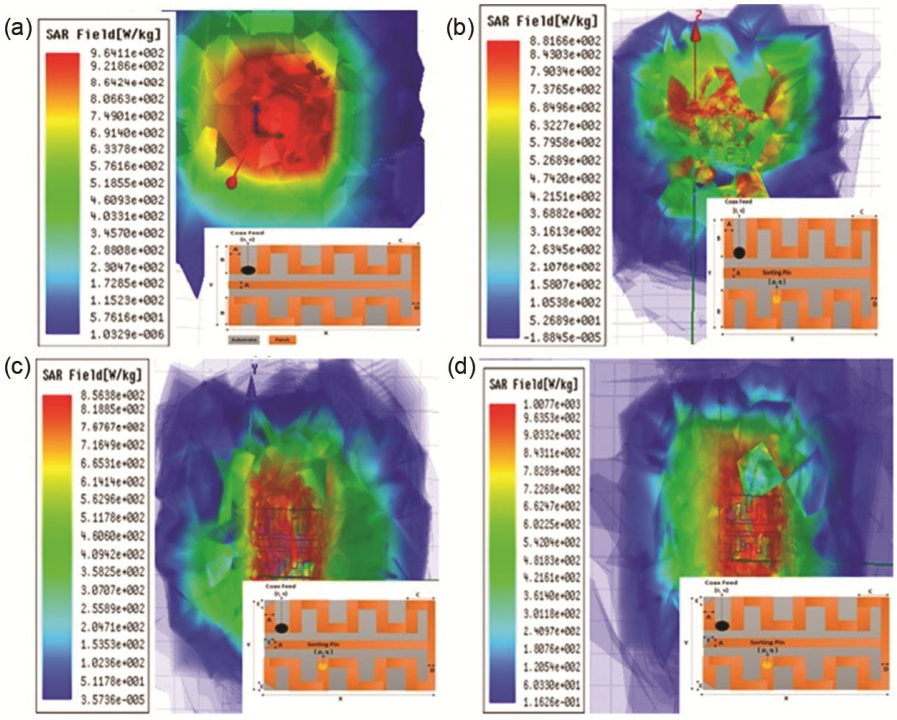


Fig. 5 — SAR Analysis (a) SAR analysis of 1st designed antenna at 1.48 GHz (b) SAR analysis of 2nd designed antenna at 2.34 GHz (c) SAR analysis of final proposed antenna with slots and shorting pin at 2.43 GHz. (d) SAR analysis of final proposed antenna with slots and shorting pin at 0.93 GHz.

establish wireless communication between BMDs and external receiver station.

4 Evaluation of SAR Distribution

As per IEEE standard safety guidelines, the human body should not absorb 1.6 w/kg or above for 1 gram of human body tissue and 2 w/kg or above for 10 grams of human body tissue[8]. So, designing an implantable antenna with this limited SAR value is our prime concern. The absorbed power by the human body and electric field is given as.

$$\text{SAR}=\int_{sample}\frac{\sigma(r)\,|\,\vec{E}(r)|^2}{\rho(r)}\,dr \qquad ...(1)$$

$$\vec{E}(\text{r})=\frac{1}{jw\epsilon\mu}[\kappa^2\vec{A}+\nabla(\nabla\cdot\overrightarrow{A})] \qquad ...(2)$$

Where $\sigma(r)$ a current density of antenna

$\vec{A}$ is the magnetic vector potential.

$\vec{E}(\text{r})$ is the electric field with the function of distance r from an antenna.

$\kappa^2\vec{A}$ is an electric field in the far-field region.

$\nabla(\nabla\cdot\overrightarrow{A})$ is an electric field in the near field region.

From equation (2) magnetic vector potential follows the spatial variation of current distribution (J).

Table 2 — SAR and Input Power of all designed antennas.

Antennas	Max SAR for 1gram	Max input power (mW) C95.1-1999(1g-avg)
1st designed antenna	964.1	1.65
2nd designed antenna	881.6	1.81
Final proposed antenna at Upper ISM Band (2.43 GHz)	856.3	1.87
Final proposed antenna at Lower ISM Band (0.95 GHz)	1007.7	1.58

Slope discontinuity of magnetic vector potential is caused by double derivative

(2) which increases the value of an electric field and hence SAR increases. SAR will be increased if an electric field is high, but this value can be reduced by having uniform current distribution along the antenna surfaces and the electric field in the near region from the antenna dominates and absorbs power. This parameter is the most important concern for patient safety. All designed antennas are having an acceptable range of SAR as demonstrated in Fig. 5.

For 1 mW input power of transmitting antenna, the maximum value of SAR would be 1600 instead of 1.6. The SAR of designed antennas with their input transmitting power is listed in Table 2. The 1st

Table 3 — Comparison Table of Proposed Antennas with Recent Research Published Work.

Ref	fr (GHz)	Volume (mm3)	Gain (dBi)	Bandwidth%	SAR (1gram)	Ground slot	Via
[9]	0.402, 0.915, 2.45	52.5	-40, -32, -22	15, 9.9, 4.2	665, 837, 759	Yes	Yes
[10]	0.92	486.4	-	11	-	No	Yes
[11]	2.46	127	-20.4	10.2	213	No	No
[12]	0.402, 2.4	642.62	-36, -27	7.4, 6.6	832, 690	No	Yes
[13]	0.402, 2.45	179.02	-30, -22	23, 21.3	-	Yes	Yes
[14]	0.915, 2.45	24	-28, -22	9.8, 8.5	971, 807	No	No
[15]	2.45	80.01	-23.1	12.8	-	Yes	Yes
[16]	2.45	91.9	-20.3	18.3	649	Yes	Yes
Final proposed antenna	0.93, 2.43	12	-21.65, -28.28	7.7, 5.5	1007.7, 856.3	No	Yes

designed antenna depicts an acceptable range of 964.1 W/KG SAR as represented in Fig. 5(a). However, due to this, transmitting power of 1st designed antenna reduces to 1.65 mW. The 2nd designed antenna has shown SAR value of 881.6 W/KG as observed from Fig. 5(b) which is lesser than the previous one. Later, for final proposed antenna, the SAR has been enhanced up to a great extent as shown in Fig. 5(c), and this proposed antenna is increased to 1.87 mW at 2.43 GHz and 1.58 mW at 0.95 GHz. The above results reaffirm the higher transmitting power, which ensures the reliability of the communication link of an implantable antenna.

At this point, it is required to compare different performance parameters of the proposed antenna design with recent works as tabulated in Table 3. It is observed that the proposed work provides the least volume of antenna which is quite significant for biomedical implantable devices, and the proposed antenna has dual ISM bands. One band can be used to transmit data during telemetry sessions and the other band can recharge the device's battery inside the human body wirelessly. The proposed antenna is also having a reduced value of SAR which ensures patient safety and reliable communication link.

5 Conclusion

A highly compact Zig-Zag-shaped implantable antenna is simulated and analyzed for biomedical application. The design shows great compactness in the size of the antenna (12 mm^3) and exhibits dual ISM bands for biomedical devices. Also, in the odd environment of the human body where absorption always degrades antenna performance, the proposed antenna has a low value of SAR (856.3) and good directive gain (-28.68 dBi) at a 2.43 GHz resonating frequency. The proposed antenna performance like reflection coefficient, gain, impedance bandwidth, SAR value, dual ISM bands have been enhanced using sorting Pin and slots technologies on the patch surface. To validate proposed antenna results, skin mimicking gel is also prepared at 2.43 GHz frequency. Proposed antenna is fabricated in microwave lab and successfully tested in skin mimicking gel solution (vitro model). With best of our knowledge, the proposed fabricated antenna has least volume among latest related research work. Good agreement between simulated and measured results has been achieved.

6 Future scope

For the future aspect, proposed antenna can be transformed to circular polarised antenna along with improved directive gain. The rectenna system can also be designed for wireless charging of embedded battery in BMDs.

References

1 Halperin D, Kohno T, Benjamin H T S, Fu K & Maisel, *IEEE Pervas Comput*, 7 (2008) 30.

2 Faisal F & Yoo H, *IEEE Trans Antennas Propag*, 67 (2019) 774.

3 Dulaimi Z A, Elwi T A & Atilla D C, *IETE J Res*, 66 (2020) 1.

4 Deng J, Li J, Zhao L & Guo L, *IEEE Antennas Wirel Propag Lett*, 16 (2017) 2270.

5 Kuster N & Balzano Q, *IEEE Trans Veh Technol*, 41 (1992) 17.

6 Gani I & Yoo H, *IEEE Microw Wireless Compon*, 26 (2016) 294.

7 Islam M ., Esselle K P, Bull D & Pilowsky P M, *IEEE Trans Microw Theory Techn*, 62 (2014) 1890.

8 Fu K, *Commun ACM*, 52 (2009) 25.

9 IEEE Standard for Safety Levels With Respect to Human Exposure to Radio Frequency Electromagnetic Fields, 3 kHz to 300 GHz, IEEE Standard C95.1(1999).

10 Liu C, Guo Y X & Xiao S, *IEEE Trans Antennas Propag*, 62 (2014) 2407.

11 Liu Y, Chen Y, Lin H & Juwono F, *IEEE Antennas Wireless Propag Lett*, 15 (2016) 1791.

12 Duan Z, Guo X, Je M & Kwong D L, *IEEE Trans Antennas Propag*, 62 (2014) 2430.
13 Ali S A & Yoo H, *IEEE Trans Antennas Propag*, 66 (2018).
14 Chow E, Yang C & Irazoqui P P, *Wireless Power Transfer*, Aalborg, Denmark, (2012) 9.
15 IEEE Standard for Safety Levels With Respect to Human Exposure to Radio Frequency Electromagnetic Fields, 3 kHz to 300 GHz, IEEE Standard C95.1, (1999)
16 Erdinc D, Ucar M H B & Adnan S, *J Turkish Chem Soc*, (2006).
17 Erdinc D, Ucar M H B & Adnan S, *J Turkish Chem Soc*, (2016) 583.

2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT) | 978-1-6654-1005-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICICICT54557.2022.9917652

Comparison of MPPT Algorithms Under Uniformly Varying Atmospheric Conditions

Chirag Kaushik
Electrical Engineering Department
Delhi Technological University
New Delhi, India
chiragkaushik201@gmail.com

Prof. Rachana Garg
Electrical Engineering Department
Delhi technological University
New Delhi, India
rachana16100@yahoo.co.in

Prof. Priya Mahajan
Electrical Engineering Department
Delhi technological University
New Delhi, India
priyamahajan@dtu.ac.in

***Abstract*— Power output from a photovoltaic (PV) array changes based on the change in ambient temperature and solar irradiation. Conventional Maximum Power Point (MPP) techniques were adopted to extract maximum power from PV array which were easy to implement but multiple drawbacks were observed related to oscillations around MPP and slower response under varying atmospheric conditions. To overcome such drawbacks of conventional algorithms, PV system requires some intelligent techniques to track the maximum power point which may help to design more efficient PV system. Authors in the present study have highlighted the comparison of multiple conventional and intelligent control techniques such as Perturb and Observe(P&O), Incremental Conductance (INC), Type 1 Fuzzy logic controller (T1-FLC) and Artificial Neural Network (ANN) based techniques. Uniform variation of solar irradiation and ambient temperature has been considered to compare the oscillations in output power waveform, power loss and fill factor of PV array. Superiority of ANN based technique has been established for all the considered factors. The work has been carried out using MATLAB/SIMULINK environment.**

Keywords—Artificial Neural network, Boost Converter, Photovoltaic systems, Maximum Power Point tracking, Fill Factor, Type 1 Fuzzy logic Controller

I. INTRODUCTION

Photovoltaics(PV) have gained great acceptance for electricity generation due to its sustainability and cleaner alternative to conventional resources. PV cell converts the photon energy from solar irradiations to electrical energy using photovoltaic effect. PV module consists of multiple PV cells connected in series. PV modules connected in parallel and series builds up PV array. These PV arrays can be used for grid integrated system or standalone system. Power output from PV array depends upon the amount of solar irradiations and the ambient temperature. PV output current changes dramatically with the variation in insolation, but voltage remains relatively constant .PV output voltage varies with variation in ambient temperature. Maximum Power point (MPP) technique works on the concept of impedance matching to trace the maximum power point of PV array so as to ensure that the maximum power is extracted from PV array. This can be achieved using chopper which is fired at optimum duty cycles using Maximum Power Point tracking(MPPT) controllers.

In literature, Villalva, M. G., and Gazoli, J. R. has proposed modelling of PV arrays [1]. D. Sinha, A. B. Das,D, K. Dhak and P. K. Sadhu has demonstrated that PV array exhibits the nonlinear characteristics of voltage and current this characteristic directly depend on the solar irradiation and some environmental factors [2]. N. Karami, N. Moubayed and R. Outbib has highlighted the theoretical comparison of multiple MPPT techniques and highlighted the general review for the same[3]. A. Anzalchi and A. Sarwat has highlighted the working of artificial neural network (ANN) based Maximum PowerPoint technique in PV system [4]. Oumnia Lagdani , Mourad Trihi and Badre Bossoufi has presented the implementation of three different types of Maximum power point tracking techniques i.e., Perturb &observe(P&O), Incremental Conductance(INC) and Type 1 fuzzy logic-based controller(T1-FLC) [5]. P.N, Jyothy & M R, Sindhu has presented the detailed theory for implementation of artificial neural network along with its performance [6]. Liu, Chun-Liang & Chen, Jing-Hsiao & Liu, Yi-Hua & Yang, Zong-Zhen has highlighted the implementation of asymmetrical fuzzy logic control based MPPT algorithm. Also, two design procedures for determinations of Universe of discourse(UOD) of membership function are also proposed[7].In the present work , the authors have established the comparison between different MPPT techniques i.e. P&O, INC, T-1 FLC and ANN based MPPT algorithms comparisons under uniformly varying atmospheric conditions. The comparisons are done on the basis of fill factor, efficiency and oscillations in power output waveforms of PV array.

This paper contains five sections. Section II introduces the PV array used in the present work. Section III gives the brief overview of designing of Boost converter. In section IV, T1-FLC is designed using Mamdani inference system. Section V highlights the trained ANN along with its performance curve for 1000 epochs. Simulation results of all the MPPT techniques are compared in section VI. Section VII concludes the paper with the concise observations gained from the proposed work.

II. PV ARRAY

Using photovoltaic effect, electricity is generated from sunlight in PV cell. PV panels are made from multiple PV cells which are connected in parallel and series combination. Bypass diodes are used for safe practical implementation of series and parallel connected solar cells which is referred as solar module. Multiple PV modules in parallel and series connection forms PV array. In the present study, authors have considered a 13.68 KW PV array consisting of 180 SunPower Modules (SPR- 76R-BLK-U) . The array consists of 4 parallel strings where each string comprises of 45 series connected modules. Specifications of the PV module under standard testing conditions are mentioned in table I[18]. Fig1. shows Power vs Voltage curves of PV array under uniformly varying solar irradiations at constant ambient temperature i.e. 25^{o}C. Fig 2 shows Power vs Voltage curves of PV array under uniformly varying ambient temperature with constant solar irradiation i.e. 1000W/m^2.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:31:54 UTC from IEEE Xplore. Restrictions apply.

TABLE I. PV MODULE SPECIFICATIONS AT STANDARD TESTING CONDITIONS

Maximum Power (P_M)	76W
Open Circuit Voltage (V_{OC})	16.2V
Short Circuit Current (I_{SC})	6.02A
Voltage at maximum Power point	13.45V
Current at maximum Power point	5.65A
Temperature Coefficient of V_{OC} (% / °C)	-0.379
Temperature Coefficient of I_{SC} (% / °C)	0.031

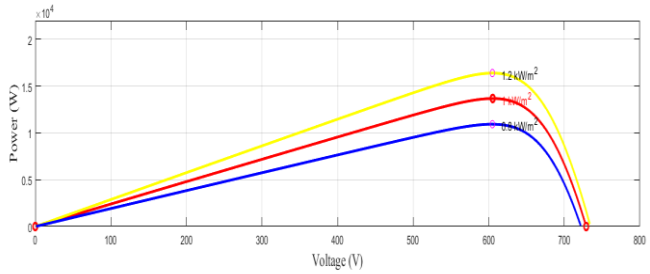


Fig 1. Power vs Voltage curve under uniformly varying solar irradiations

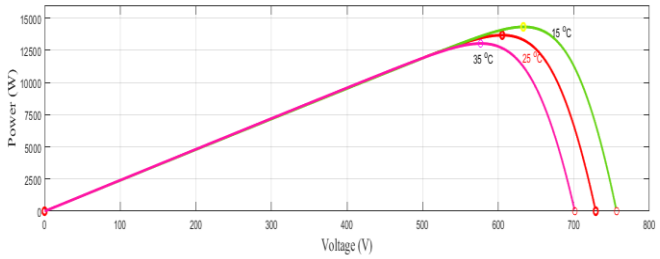


Fig 2.Power vs Voltage curve under uniformly varying ambient temperature

III. BOOST CONVERTER

The impedance matching is done using DC-DC converter to produce maximum power from PV array. Boost converter has been considered in the present study which transfers power from PV array as a certain voltage to DC load at higher voltage. In the present work, the boost converter is fired by the MPPT controller to achieve impedance matching. Circuit diagram of the boost converter along with its interfacing with PV array, MPPT controller and load is shown in Fig 1

The modelling of Boost Converter is done using following equations:

$$L = \frac{V_s \alpha}{\Delta I f} \quad (1)$$

$$C = \frac{I \circ \alpha}{f \Delta V_c} \quad (2)$$

$$V_o = \frac{1}{1-\alpha} V_s \quad (3)$$

where V_s is the input supply voltage at MPP; α is the duty cycle of converter; f is the frequency at which converter is working; ΔI output current ripple which is considered 5% of the input current at MPP and α (duty cycle) is taken as 0.5 so as to obtain the inductor value for maximum ripple. ΔVc is ripple in load voltage which is considered less than 5% of load voltage Vo.

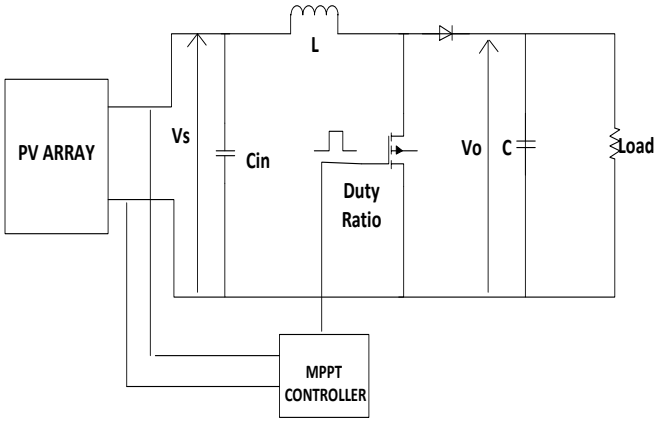


Fig. 3 Schematic diagram for proposed standalone PV system

IV. TYPE 1 FUZZY LOGIC CONTROLLER

Fuzzy Logic controller(FLC) works on the fuzzy logic theory [8]. Fuzzy logic theory helps reduce the effects of uncertainties. Fuzzy controller comprises of three processes: fuzzification, rule processing and defuzzification. Fuzzifier block performs the fuzzification process in which the crisp input is converted to linguistic variables based on membership functions, authors have used five fuzzy levels: NB , NS ,ZE, PS and PB . These are further processed by rule processing unit where desired output is generated depending upon the rule base. Finally, crisp output is generated using defuzzifier A fuzzy block diagram is shown in fig. 4.

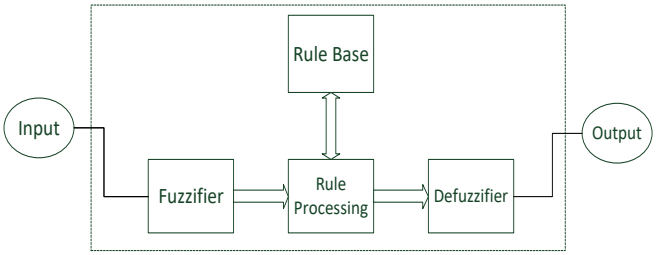


Fig. 4 Structure of Fuzzy logic Controller

In the proposed system, the inputs of FLC is considered as change in voltage (dV) of PV array and change in PV array output power(dP). Fig 5 and fig 6 highlights the membership functions of the inputs of FLC considered in the present work. The required change in duty cycle of boost converter is considered as the output of the FLC as shown in fig 7. The output from FLC can be found out using the rule base shown in Table 2.

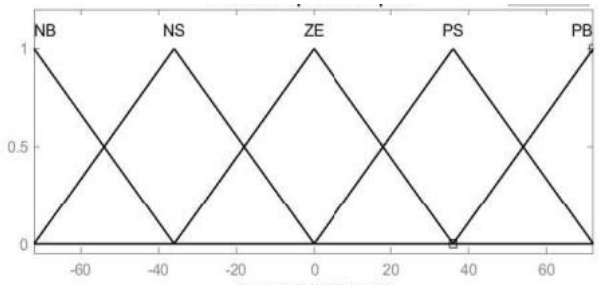


Fig 5. Membership functions for change in PV array output power

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:31:54 UTC from IEEE Xplore. Restrictions apply.

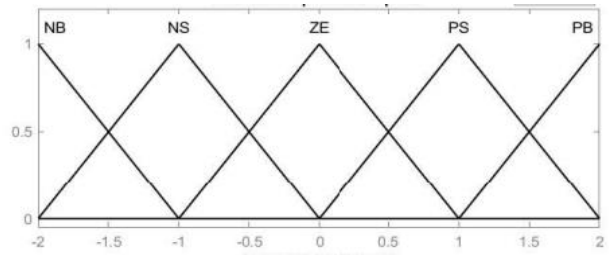


Fig 6. Membership functions for change in PV array voltage

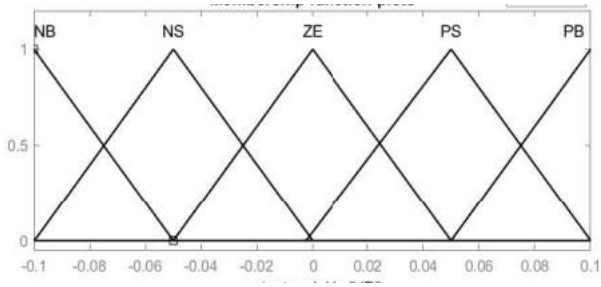


Fig 7. Membership functions for change in duty Cycle

TABLE II. FUZZY RULE BASE

dP \ dV	NB	NS	ZE	PS	PB
NB	NS	NB	PB	PB	PS
NS	ZE	NS	PS	PS	ZE
ZE	ZE	ZE	ZE	ZE	ZE
PS	ZE	PS	NS	NS	ZE
PB	PS	PB	NB	NB	NS

V. ARTIFICIAL NEURAL NETWORK

Artificial neural network (ANN) is a type of machine learning model inspired by human brain, which are generally used to perform tasks like clustering, classification and pattern recognition. ANN comprises of interconnected "neurons" which send signals to each other. The connections are assigned numeric weights which are tuned based on experience gained from training the network so as to make the neural network adaptive to inputs and capable of learning. There are multiple types of neural network architecture. In the present study, authors have considered multilayer perceptron model with inputs as ambient temperature and solar irradiance and the output is the reference voltage at which the PV array exhibits maximum power. Weighted linear sum is considered as basis function for hidden layer and output layer. Transig is considered as activation function for hidden layer whereas output layer has linear activation function. The algorithm used to train the ANN is Levenberg-Marquardt which requires more memory but takes less time compared to other algorithms. Mean square error is considered as the parameter to determine the performance of the proposed ANN, training of the ANN stops when there is no improvement in mean square error The training data is generated using MATLAB script program which produces matrix for solar irradiation varying from 0 to1000 and ambient temperature varying from 15ºC to 35ºC. Among 1000 data sets 700 data sets are used for training, 150data sets for validating and 150 for testing. The hidden layer consists of 10 neurons. The neural network was created using neural network wizard provided by MATLAB/Simulink. The error histogram is shown in fig. 8. Also, the performance curve of trained ANN is shown in fig.9 which shows Mean squared error of order 10^{-8} was attained after 1000 epochs

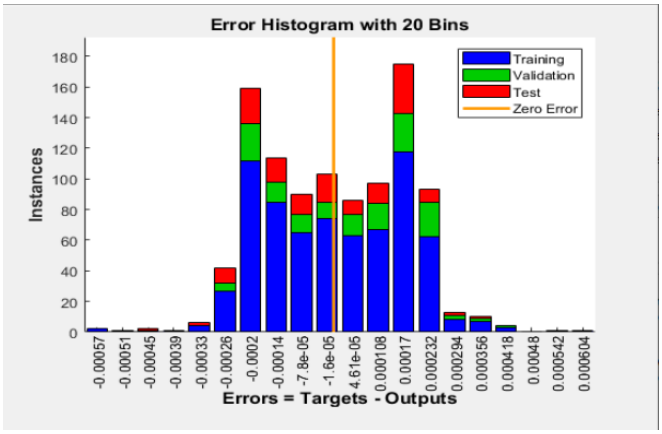


Fig.8 Error histogram of Artificial neural Network

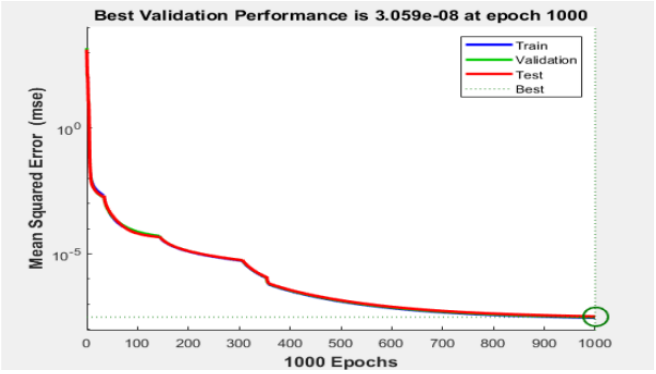


Fig 9 Performance Curve Of the Artificial Neural Network

VI. SIMULATION RESULTS AND DISCUSSIONS

A 13.68 kW standalone PV system is tested under uniformly varying atmospheric conditions using a simulation model in Matlab/Simulink. The dependence of PV array on ambient temperature and solar irradiation is discussed in this section. PV array is tested with uniformly varying solar irradiations at constant ambient temperature and uniformly varying ambient temperature at constant solar irradiations. The solar irradiance is varied at constant ambient temperature from 800(W/ m^2) to 1200(W/ m^2) in steps of 200(W/ m^2). with time instances as 0.5 sec and 1 sec. The ambient temperature is varied at constant solar irradiations from 35ºC to 15 ºC in steps of 5 ºC. with time instances as 0.5 sec and 1 sec.

A. Case 1(Uniformly varying solar Irradiation and constant ambient temperature i.e. 25℃):

Fig 10 highlights the extracted power waveforms using P&O, INC, T1-FLC and ANN based MPPT controllers. It is quite evident that Power output from PV array increases with increase in solar irradiations at constant ambient temperature. Also, it is noted that power output from PV array reduces with reduction in solar irradiations at constant ambient temperature.

ANN based MPPT controller outperformed T1-FLC, INC and P&O based MPPT controller under uniformly varying solar irradiations at constant ambient temperature. When the PV array was operating at 800(W/m^2),it has extracted

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:31:54 UTC from IEEE Xplore. Restrictions apply.

maximum power of 10930.8W from PV array with ripple of around 2W. When the PV array was operating at 1000(W/m^2), it has extracted maximum power of 13677.33W from PV array with ripple of around 1.57W. When the PV array was operating at 1200(W/m^2), it has extracted maximum power from PV array i.e. 16394.3W with ripple of around 2.84W.

From table III, it has been found out that T1-FLC based MPPT controller has performed better than P&O and INC based MPPT controllers as T1-FLC generated ripples of around 8.14W whereas INC based MPPT controller has generated ripples of around 26.3W and P&O based MPPT controller has generated ripples of around 54W. It can also be noted that lesser ripples were produced under standard test conditions as compared to other testing conditions mentioned above.

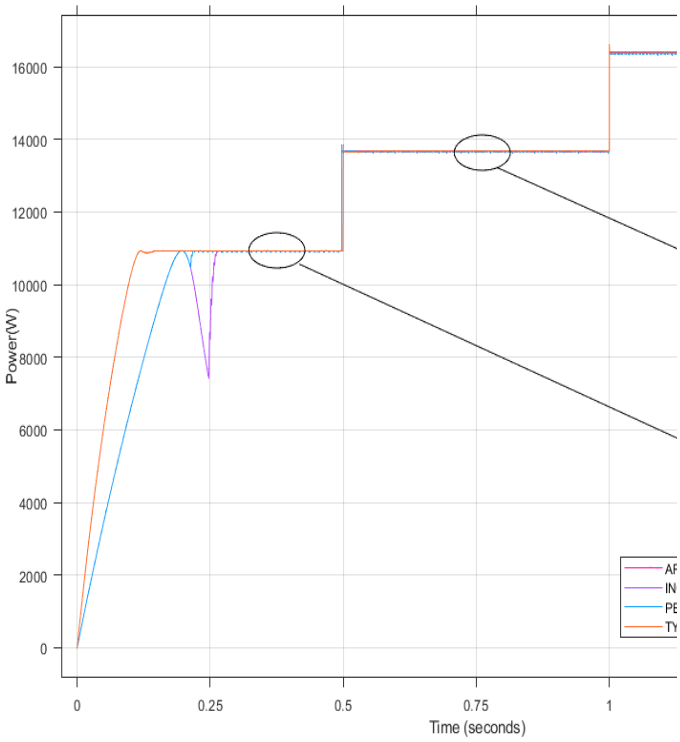


Fig.10: Power vs Time plot of output power of PV array for various MPPT techniques under uniformly varying solar irradiation

TABLE III. PV ARRAY OUTPUT POWER UNDER UNIFORMLY VARYING SOLAR IRRADIATIONS

Testing conditions		**PV Array Power Output**							
Ambient Temperature (ºC)	***Irradiance (W/m^2)***	***P&O***		***INC***		***T1-FLC***		***ANN***	
		Power Extracted (W)	***Ripple (W)***	***Power Extracted (W)***	***Ripple (W)***	***Power Extracted (W)***	***Ripple (W)***	***Power Extracted (W)***	***Ripple (W)***
25	800	10907.9	48	10917.65	28.7	10928.95	9.1	10930.8	2
25	1000	13653.7	50	13667.3	21.7	13674.3	7.26	13677.33	1.57
25	1200	16357.7	72.8	16375.7	39.7	16385.85	13.9	16394.3	2.84

B. Case 2 (Uniformly varying ambient Temperature and constant solar irradiation):

Fig 11 shows the extracted power waveforms using P&O, INC, T1-FLC and ANN based MPPT controllers. It is quite evident that Power output from PV array increases with decrease in ambient temperature at constant solar irradiations. Also, it is noted that power output from PV array reduces with increase in ambient temperature at constant solar irradiations

ANN based MPPT controller outperformed T1-FLC, INC and P&O based MPPT controller under uniformly varying solar irradiations at constant ambient temperature. When the PV array was operating at 35ºC, it has extracted maximum

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:31:54 UTC from IEEE Xplore. Restrictions apply.

power of 13030.8W from PV array with ripple of around 1.23W. When the PV array was operating at 25ºC it has extracted maximum power of 13677.33W from PV array with ripple of around 1.57W. When the PV array was operating at 15ºC, it has extracted maximum power from PV array i.e., 14321.5W with ripple of around 1.82W.

From table IV, it can be seen that T1-FLC based MPPT controller has performed better than P&O and INC based MPPT controllers as it generated ripples of around 6.2W whereas INC based MPPT controller has generated ripples of around 22.6W and P&O based MPPT controller has generated ripples of around 52.3W.

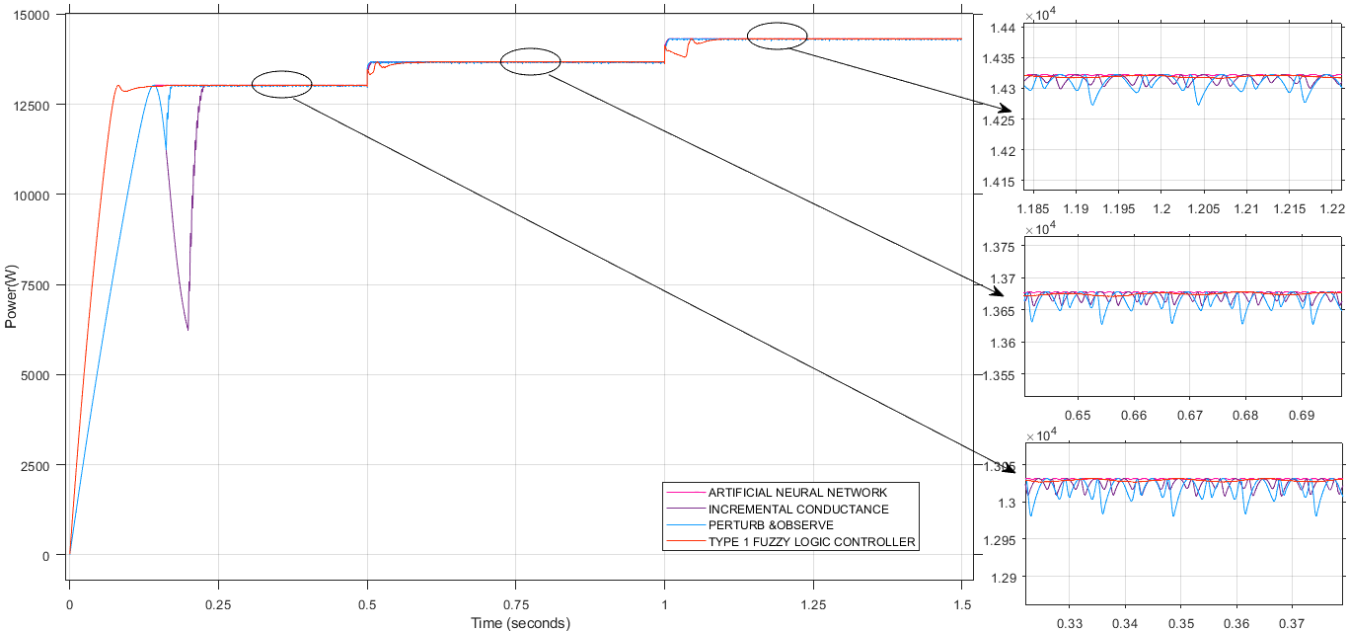


Fig.11: Power vs Time plot of output power of PV array for various MPPT techniques under uniformly varying ambient temperature

TABLE IV. PV ARRAY OUTPUT POWER UNDER UNIFORMLY VARYING AMBIENT TEMPERATURE

Testing conditions		**PV Array Power Output**							
Temperature (ºC)	***Irradiance (W/m²)***	***P&O***		***INC***		***T1-FLC***		***ANN***	
		Power Extracted (W)	***Ripple(W)***	***Power Extracted (W)***	***Ripple (W)***	***Power Extracted (W)***	***Ripple (W)***	***Power Extracted (W)***	***Ripple (W)***
35	1000	13006.15	50.5	13019.65	23.5	13028.23	6	13030.8	1.23
25	1000	13653.25	49.7	13667.3	21.7	13674.3	7.26	13677.33	1.57
15	1000	14294	56.6	14311	22.6	14319.92	5.34	14321.5	1.82

Fill Factor (FF): It is defined as the percentage of maximum Power output from PV array (P_m) to the product of open circuit voltage (V_{oc}) and short circuit current (I_{sc}) of the PV array at given atmospheric conditions. As the fill factor moves closer to unity, the more efficient is the system. Table V and VI highlights the fill factor of all the MPPT algorithms under uniformly varying atmospheric conditions.

$$FF = \frac{Pm}{Voc*Isc} *100 \quad (4)$$

TABLE V. FILL FACTOR UNDER UNIFORMLY VARYING SOLAR IRRADIATIONS

Testing Conditions		P&O	INC	T1-FLC	ANN
Temperature(C)	Irradiance (W/ m²)				
25	1000	77.77	77.82	77.88	77.91
25	800	78.45	78.54	78.6	78.62
25	1200	77.16	77.25	77.29	77.34

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:31:54 UTC from IEEE Xplore. Restrictions apply.

TABLE VI. FILL FACTOR UNDER UNIFORMLY VARYING AMBIENT TEMPERATURE

Testing Conditions		P&O	INC	T1-FLC	ANN
Temperature(C)	Irradiance (W/ m^2)				
25	1000	77.77	77.82	77.88	77.91
15	1000	78.21	78.3	78.34	78.37
35	1000	76.77	76.85	76.9	76.92

Power Loss: Power loss is calculated as difference between theoretical maximum power output from PV array and maximum power output from PV array.

$$\text{Power Loss} = P_{\text{theo. max}} - P_{\text{o/p pv}} \quad (5)$$

From table V and VI it is found out that ANN based MPPT algorithm has highest fill factor while P &O has lowest fill factor out of all the MPPT algorithms mentioned above. It has been observed that fill factor is directly proportional to ambient temperature and inversely proportional to solar irradiations

TABLE VII. POWER LOSS UNDER UNIFORMLY VARYING SOLAR IRRADIATIONS

Testing Conditions		P&O	INC	T1-FLC	ANN
Temperature(C)	Irradiance (W/ m^2)				
25	800	24	14.3	4.6	1
25	1000	24.8	10.85	3.63	0.78
25	1200	36.4	19.85	6.95	1.42

TABLE VIII. POWER LOSS UNDER UNIFORMLY VARYING AMBIENT TEMPERATURE

Testing Conditions		P&O	INC	T1-FLC	ANN
Temperature(C)	Irradiance (W/ m^2)				
35	1000	25.25	11.75	3	0.6
25	1000	24.8	10.85	3.63	0.78
15	1000	28.3	11.3	2.67	0.91

From table VII and VIII it has been observed that ANN based MPPT algorithm is the most efficient with power loss of around 0.91W while P &O is least efficient MPPT algorithm with power loss of around 27.2W.

VII. CONCLUSION

In the present study, all the above mentioned MPPT algorithms are tested under two cases i.e., constant ambient temperature with uniformly varying solar irradiations; and constant solar irradiation with uniformly varying ambient temperature. It has been observed that intelligent controllers performs better than con ANN based MPPT algorithm has outperformed all other MPPT algorithm. On the other hand, it can also be noted that T1-FLC based MPPT algorithm has performed much better than P&O and INC MPPT algorithms. Further, the comparisons are done on the basis of fill factor, efficiency and Mean output power from PV array. It was seen ANN was the most efficient MPPT algorithm out of all other algorithms mentioned above and P&O was least efficient. The fill factor for ANN was much better than other algorithms. It can be concluded that intelligent MPPT techniques has performed much better than conventional MPPT techniques.

REFERENCES

[1] Villalva, M. G., and Gazoli, J. R. “Comprehensive approach to modelling and simulation of photovoltaic arrays,” IEEE Trans. Power Electron., Vol. 24, No. 5, pp. 1198–1208, 2009.

[2] D. Sinha, A. B. Das,D, K. Dhak, P. K. Sadhu “Equivalent circuit configuration for solar PV cell” in 1st International Conference on Non-Conventional Energy (ICONCE), 2014

[3] N. Karami, N. Moubayed and R. Outbib, “General review and classification of different MPPT Techniques,” Renewable and Sustainable Energy Reviews, Volume 68, Part 1, February 2017, Pages 1-18.

[4] A. Anzalchi and A. Sarwat, "Artificial neural network based Duty Cycle estimation for maximum Power Point tracking in Photovoltaic systems," SoutheastCon 2015, 2015, pp. 1-5, doi: 10.1109/SECON.2015.7132988.

[5] Lagdani, Oumnia et al. “PV array connected to the grid with the implementation of MPPT algorithms (INC, P&O and FL Method).” International Journal of Power Electronics and Drive Systems 10 (2019): 2084-2095.

[6] P.N, Jyothy & M R, Sindhu. (2018). An Artificial Neural Network based MPPT Algorithm For Solar PV System. 375-380. 10.1109/ICEES.2018.8443277.

[7] Liu, Chun-Liang & Chen, Jing-Hsiao & Liu, Yi-Hua & Yang, Zong-Zhen. (2014). An Asymmetrical Fuzzy-Logic-Control-Based MPPT Algorithm for Photovoltaic Systems. Energies. 7. 2177-2193. 10.3390/en7042177.

[8] Zadeh, Lotfi Asker. "Fuzzy sets as a basis for a theory of possibility." Fuzzy sets and systems 1.1 (1978): 3-28.

[9] Liu, Ping & Diao, Wan. (2014). Optimization Design of Photovoltaic System MPPT Controller. Applied Mechanics and Materials. 571-572. 940-943. 10.4028/www.scientific.net/AMM.571-572.940.

[10] Singh, Hawa & Shahid, Muhammad. (2016). Soft Computing Techniques FUZZY and ANN based MPPT for Grid Tied PV. International Journal of Engineering Research and. V5. 10.17577/IJERTV5IS050620

[11] Nedumgatt JJ, Jayakrishnan KB, Umashankar S, Vijayakumar D, Kothari DP. Perturb and observe MPPT algorithm for solar PV systems-modeling and simulation. In: Proc - 2011 Annu IEEE India Conf eng sustain solut INDICON-2011, vol. 19. 2011. p. 1–6.

[12] hmad J. A fractional open circuit voltage based maximum power point tracker for photovoltaic arrays. In: ICSTE 2010-2010 2nd int conf softw technol eng proc, vol. 1. 2010. p. V1-247-V1-250.

[13] Verma P, Garg R, Mahajan P. Asymmetrical fuzzy logic control based MPPT algorithm for stand-alone PV system under partially shaded conditions. Int J Sci Technol 2010.

[14] tsoy O, Colak S. Adaptive estimator design for unstable output error systems: A test problem and traditional system identification based analysis 2015;229:902–16.

[15] Arvind Kumar, Manoj Kumar, Dattatraya H. Nagaraj, Amanpreet Singh and Jayanthi Prattapati, “Comparative Study of IC and Perturb and Observe Method of MPPT Algorithm for Grid Connected PV Module”. World Academy of Science, Engineering and Technology International Journal of Electrical and Computer Engineering, Vol. 8, No. 7, pp. 1225-1233, 2014.

[16]M . Lokanadham and K.Vijaya Bhaskar, “Incremental Conductance Based Maximum Power Point Tracking (MPPT) for Photovoltaic System”. International Journal of Engineering Research and Applications (IJERA), Vol. 2, Issue 2, pp.1420-1424, Mar-Apr 2012

[17] A .S. Swathy and R. Archana, “Maximum Power Point Tracking Using Modified Incremental Conductance for Solar Photovoltaic System”. International Journal of Engineering and Innovative Technology (IJEIT), Vol 3, Issue 2, pp. 333-337, August 2013.

[18] http://www.solarhub.com/solarhub_products/780-SPR-76R-BLK-U-SunPower

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:31:54 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3215264

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2022.Doi Number

Deep Learning Models Performance Evaluations for Remote Sensed Image Classification

Abebaw Alem[1, 2] and Shailender Kumar[1]
[1]Department of Computer Science and Engineering, Delhi Technological University, Delhi-110042, India
[2]Faculty Member of IT department, Debre Tabor University, Debre Tabor, Ethiopia

Corresponding Author: Abebaw Alem (Email: abebale@dtu.edu.et)

This work was supported by the Ethiopian Ministry of Education (MoE) by providing a scholarship with a monthly stipend financial support.

ABSTRACT Deep learning-based land cover and land use (LCLU) classification systems are a significant aspiration for remote sensing communities. In nature, remote sensing images have various properties that need to be analyzed. Analyzing and interpreting image properties is difficult due to the image nature, the sensor technology's capability, and other determinant variables such as seasons and weather conditions. The problem is essential for environmental monitoring, agricultural decision-making, and urban planning if it can be supported in deep learning systems. Therefore, deep learning approaches are proposed to quickly analyze and interpret the remote sensing image to classify the LCLU. The deep learning methods could be designed starting from scratch or using the pre-trained networks. However, comparisons of the deep learning method developed from scratch and trained on the pre-trained networks are scarce. Thus, we proposed evaluating and comparing the deep learning models convolutional neural network feature extractor (CNN-FE) by developing from scratch, transfer learning, and fine-tuning for LCLU classification system using remote sensed images. This paper compares and analyses the deep learning remote sensed image classification algorithms by designing CNN-FE, TL, and fine-tuning deep learning models as examples. After developing and training each deep learning model on the UCM dataset, we evaluated and compared their performances using the performance measurement metrics accuracy, precision, recall, f1-score, and confusion matrix. The proposed deep learning algorithms can adapt and learn the features of the remote sensing images, and the TL and fine-tuning classification performances are significantly improved. Thus, this paper found that the fine-tuned deep learning model has achieved profound accuracy performance results in the UCM dataset with the efficient time used for training the models.

INDEX TERMS Convolutional neural network**,** deep learning, fine-tuning, performance comparisons, remote sensed image classification, transfer learning.

I. INTRODUCTIONS

The land cover and land use (LCLU) classification learning system is essential for environmental monitoring, agricultural decision-making, and urban planning in the contemporary dynamic world. The LCLU classification using remote sensed (RS) images is a critical issue in managing natural resources and human-made activities on the natural phenomenon in the earth's environment. RS image classification is the recently focused area for the RS societies in the computer vision trends and image processing research areas. From time to time, the world's population is increasing dramatically, and the demand for land use is increasing. A learning system could be applied to the domain to utilize this land properly.

Thus, the LCLU classification is the recent hot, challenging task in RS [1]–[4]. With advanced sensor technologies, RS images are the satellite data collected from the earth's environment. The deep learning (DL) method could be applied to solve the challenge.

The DL is the recent specialized machine learning (ML) approach that could automatically extract features of the image for large datasets with admirable performance improvements. Thus, DL is the recently focused research area applied in various domains, such as classification [1], [5]–[8], recognitions [9], and object detections [10]. It is also potentially challenging in many other domains [11].

DL techniques proposed in this paper are convolutional neural network (CNN), transfer learning (TL), and fine-tuning, which make the classification task more attractive. CNN is one of computer vision's most common DL methods [18] for feature extractions and LCLU modeling using RS images. The CNN is a feedforward and a backward neural network consisting of convolutional calculations and deep structures. Therefore, CNN models have powerful feature extraction capability for classification performance improvement in RS images [12].

Nevertheless, the DL algorithms such as CNN require a large amount of data and very high computational power to train the classification models[13] from scratch. Whereas TL [14] and fine-tuning [4], [15] can solve the classification problem in smaller dataset training samples and less training time. Thus, the main problem of the DL models is training the DL models using deep CNN from scratch requires a large dataset and takes longer to train the model. To solve such DL problems, we proposed the TL and fine-tuning approaches, and we compared their performances with the scratch development convolutional neural network feature extractor (CNN-FE) model.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3215264

TL is another recent DL technique used to train the DL model by reusing pre-trained networks. TL and fine-tuning are used for smaller datasets and can be designed from the pre-trained networks on their top fully connected layer to reuse the features. The training time in TL and fine-tuning could be much less than that of deep CNN models trained from scratch. Therefore, TL could solve the problems of scratch development of DL models by training the models in less time with relatively small datasets by freezing the pre-trained network.

TL adopts the features from the pre-trained network to train the new models. Moreover, fine-tuning is a DL technique used to train the model by unfreezing the pre-trained networks. This technique is vital in increasing the performance of the model. The TL adopts the properties of the pre-trained layers, excluding the last fully connected layer, i.e., the dense layer is replaced by our classifier with a number of neurons is 21, and the activation function is softmax.

This paper designed and evaluated the DL models CNN, TL, and fine-tuning. The CNN has been developed from scratch with four CNN blocks. Using Keras applications, the deep CNN-based TL and fine-tuning models have been developed on the pre-trained model EffificientNetB7 [16].

In recent studies, very few studies have been conducted to compare the DL models developed from scratch and reused the pre-trained network to evaluate their capable performances. For instance, [17] have applied the TL and fine-tuning methods on the ResNet50 pre-trained network and compared their performances with other pre-trained based networks in scene image classification. However, the scratch development models' evaluation and comparison with pre-trained development models have not been widely researched. For designing the TL and fine-tuning model, we preferred the recent pre-trained network, the EffificientNetB7, which was trained on the "ImageNet" large dataset. Recently, [16] achieved 84.4% top-1 accuracy of state-of-the-art EffificientNetB7 on the "ImageNet". According to [16], eight scaling-up series of EffificientNet pre-trained models from EffificientNetB0-EffificientNetB7 were designed on the larger dataset called "ImageNet". The performance of each successive version has been improved.

According to [18], who have applied EffificientNetB3, the large versions of EffificientNet models achieve better accuracy than the smaller ones. Thus, we proposed the EffificientNetB7 pre-trained network to design TL and fine-tuning models in the domain of LCLU classification using RS images to evaluate their performances and compare them with the CNN-FE developed from scratch. We selected the UCM dataset to assess and compare the DL models.

Therefore, this paper aims to design the DL models and evaluate their performance with various performance measurement metrics. We contributed the following main contributions accordingly.

- We developed the CNN model from scratch and compared its performance with the deep TL and fine-tuning models for LCLU classification using the RS UCM dataset.
- We applied the recent advanced EfficientNetB7 pre-trained network to design TL and fine-tune DL models to LCLU classification in RS images.
- We evaluated the models, compared their performances using different performance evaluation metrics, and concluded that the fine-tuning model improved performance with efficient training time.

II. MATERIALS AND PROPOSED METHODS

A. DATASETS AND TOOLS

We used the publicly available University of California Merced (UCM) dataset for modeling the CNN, CNN-based TL, and fine-tuning. The UCM dataset is an LCLU data set collected from the earth, labeled manually, and introduced by [19] at the University of California Merced. It contains twenty-one classes. Each class includes 100 images with 256 × 256 pixels and a spatial resolution of about 30 centimeters per pixel. However, the UCM dataset is inconsistent as about 44 images have different pixel shapes. This dataset is available at http://weegee.vision.ucmerced.edu/datasets/landuse.html.

The python high-level computer language as a tool is used. Python is a flexible and easily understandable tool capable of many interactive libraries to build the DL model. Tensorflow and Keras are also other DL tools used with python.

B. PROPOSED DL METHODS

In the previous works, the DL method has been studied in the classification problems in RS images of various datasets. However, evaluating and comparing the DL developed from scratch and on the pre-trained network has not been investigated widely. Further investigations are still needed to design and assess the current DL techniques for LCLU classification using the RS datasets.

Thus, to evaluate and compare the performances of different DL models applied for LCLU classification problems in the UCM RS dataset, we designed the CNN model from scratch and the TL and fine-tuning models on EfficientNet pre-trained network. The EfficientNet was trained on the large-size dataset "ImageNet" images. ImageNet is the most significant benchmark dataset introduced by [20] for designing DL models.

The DL hyperparameters influenced on CNNs performances [21]. For instance, according to [22], using different dropout values produced different performance results. We also showed that the dropout value (0.25) generated an accuracy of 84.76 %, which is different from our previous work with the dropout value (0.50) caused an accuracy of 89.76. Therefore, by considering their effects, we set the same hyperparameters for all the three DL models on the given dataset to evaluate the models' performances, as shown in Table I.

1) THE CONVOLUTIONAL NEURAL NETWORK (CNN) ALGORITHM

The CNN algorithm is the most critical DL technique that could extract and automatically learn features from the data. Features are newly extracted and learned weights of pixels of the image in the new value (usual in reduced) from the input images. The CNN method consists of several series of connected layers. These layers shared weights throughout the process, i.e., from the start

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3215264

to the end layer (classifier), as depicted in Figure 1. This process makes the model's feature map at the entire layers and the class prediction at the output layer. The feature map of the model can be built up with pixel-wise multiplication of the input image pixel and the provided weight or kernel pixels with learnable parameters [23], [24].

The CNNs can be capable of spatial feature representations for RS image classifications using the convolution technique in the form of pixels [25]. This convolution process updates weights with each layer's provided non-linear activation function. The input data types and weight calculations in the convolution method make the CNNs different from other conventional ML approaches [26].

In DL model training, relu and softmax non-linear activation functions are the most relevant function to update the weights in the convolution process. We used the relu at the entire convolutional layers to activate the weights in each convolution process and the softmax at the output layer since it is reliable for multiclass classification problems. The softmax function is a feature classifier based on a probability score for each class.

In the convolution process, the number of parameters (params) could be calculated using equations (1) and (2). The total parameter numbers of the model are the summations of the computed results from Conv2D and Dense layers. We designed the CNN-FE model with four Conv2D layers that calculate the number of parameters for those layers in the same norm by (1) and two dens layers (2). However, the calculation formula of dense parameters differs from Conv2D as equated in (2). The number 1 means the bias associated with each filter for learning.

$$\mathrm{C2DP\#} = \#\mathrm{OC} * (\#\mathrm{OC} * \mathrm{FH} * \mathrm{FW} + 1) \quad (1)$$

$$\mathrm{DP\#} = \#\mathrm{OC} * (\#\mathrm{IC} + 1) \quad (2)$$

Where, C2DP# = number of convolutional parameters, #OC = number of output channels, DP# = number of dense parameters, #IC = number of input channels, FH= filter size of height, and FW= filter size of width

We could get a total calculated parameter number according to (1) and (2). However, the number of parameters for all MaxPooling2D and Flatten layers is zero because these layers do not learn anything from weights/filters or the built model. As a result, 1.68 million parameters were found and learned in the CNN-FE model, while 18.88 million parameters were found and learned in both TL and fine-tuning models.

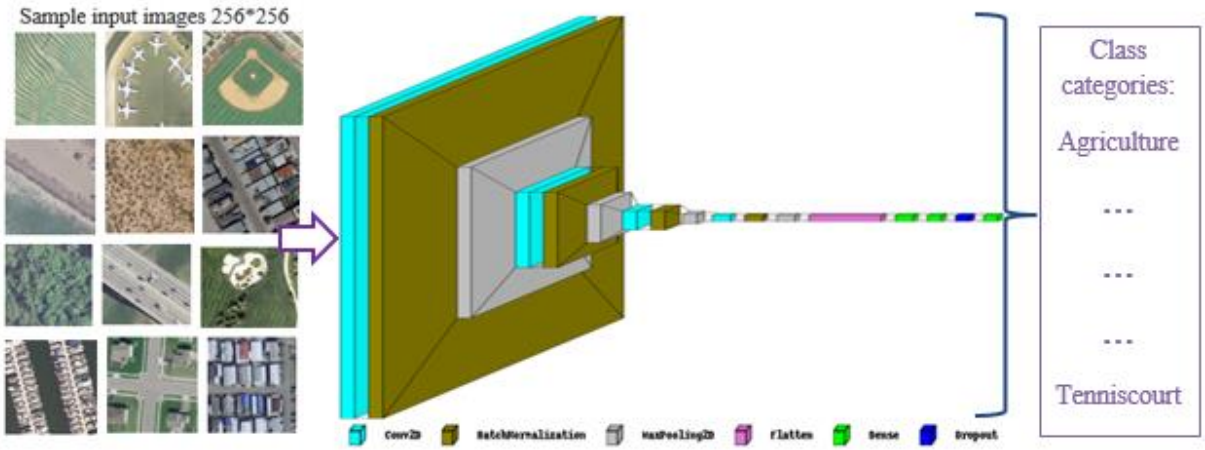


FIGURE 1. Layers of CNN-FE model with the input sample images

2) THE TRANSFER LEARNING (TL) METHOD

TL is a method of training the DL model by replacing the input layer with an image embedding as the EfficientNet transfers the knowledge learned from the much larger dataset called "ImageNet" to our classification problem. The TL has been trained by making the layers in EfficientNet on ImageNet images non-trainable (pre_trained_model.trainable = False). We trained only the last flatten (1D vector form) and two dense layers, including relu and softmax activation functions and dropout on the 21 LCLU RS UCM dataset classes. Therefore, the classification head with dense layers can be appended for manipulating our new classification problem. TL is an efficient, reliable DL technique used to propose various domains, especially the image classification problem in this paper. Recently, TL has been applied for LCLU in RS image classifications [27], [28].TL is used to train DL models in efficient processing time with improved performances [29], [30]. However, deep TL is used for limited dataset training samples, while deep CNN from scratch is used for large dataset training samples. Therefore, we applied the TL model in this paper to compare its performance with other DL techniques for LCLU classification in RS images.

3) THE FINE-TUNING ON EFFICIENTNET

Fine-tuning is a DL technique used to train a model by allowing and adapting the EfficientNet pre-trained layers on ImageNet large dataset to be trainable (pre_trained_model.trainable = True). EfficientNet is the recent advanced CNN-based network that could be applied in classification tasks on ImageNet. To get

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3215264

the improved performance, EfficientNet has been fine-tuned, and the final fully connected layer is treated as the output classifier layer as we did in TL, except the layers are allowed to be trained. As stated by [4] and [15], fine-tuning a pre-trained network is the optimal solution for a limited number of training samples. The EfficientNet pre-trained network was introduced by [16] for rethinking model scaling for CNN. We selected the EfficientNet pre-trained network as it is the most recent and advanced network, which has not been applied yet to the CLCU classification domain.

Fine-tuning the EfficientNet pre-trained network has been trained on the last three fully connected layers on the ImageNet. The final fully connected layers include a flatten layer that transforms the input image into vector form, two dense layers, dropout, relu, and activation functions. Accordingly, the pre-trained weights are used randomly as initial weights for our fine-tuning neural network. Fine-tuning is used to compare the results of the fully connected layer and convolutional layer. Thus, we proposed fine-tuning technique to compare its performance with the convolutional layer-based CNN-FE model and the fully connected layer-based TL model.

TABLE I
THE DL HYPERPARAMETERS SETTINGS FOR TRAINING THE DATASETS

Hyperparameters	Values
Optimizers	Adam
Learning rate	0.001
Batch size	64
Epochs	100
Loss function	Cross-entropy
Activation functions	Relu, softmax
Dropout	0.25

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

The experiments were executed in Intel Core i3-4000M CPU 2.40 GHz RAM= 4GB laptop personal computer integrated with Colaboratory in its Tesla K80 GPU. Keras and Tensorflow, open-source DL software packages, were used for this experiment in python. In addition, we used the scikit-learn statistical package for analyzing the performance measurement metrics.

A. EXPERIMENTAL SETTING AND RESULTS

We used the UCM dataset for experiments to design and evaluate the DL models for LCLU classification problems. We split each dataset to train, validate and test samples in 60%, 20%, and 20%, respectively. We also set the DL hyperparameters as indicated in Table I. Then, we trained the models, validated them with the validation dataset during training, and evaluated the performances with the test dataset.

After the experimental parameters setting, we trained and evaluated the model during and after the experiments with validation and test datasets, respectively. We evaluated the models using accuracy, precision, recall, f1-score, and confusion matrix (CM) metrics. CM measures the class performance, whether classified correctly or incorrectly in rows-column intersections. In addition to the accuracy, we used the categorical-cross-entropy loss function to calculate the errors. The training and validation losses or mistakes are expected to decrease in epoch increments, as shown in Figure 2, Figure 3, and Figure 4 (on the **right**).

Therefore, we evaluated the models with 420 test or support images, as shown in Table II, Table III, and Table IV, using the UCM dataset. The UCM is an imbalanced RS dataset. The accuracy performance metric, the percentage of correctly classified images, could not be suitable for the imbalanced dataset. Thus, each class performance is evaluated using precision, recall, f1-score, and CM metrics in addition to the accuracy metric. The f1-score is the harmonic mean of precision and recall metrics, and it generalized the performance of each class and the average performance of the DL models built. If both precision and recall have the best performance result in a category, then the f1-score has the best performance result in that class. Whereas, if either precision or recall has a 0-performance result, then the f1-score has 0 performance which is nothing the model is predicting.

Accordingly, the categories that have scored best (100%) in the f1-score metric are agricultural and chaparral in CNN-FE (Table II); chaparral, parkinglot, and storagetanks in TL model (Table III); and agricultural, airplane, chaparral, freeway, and runway in a fine-tuning model (Table IV), respectively.

TABLE II
CNN-FE CLASSIFICATION PERFORMANCES IN PRECISION, RECALL, AND F1-SCORE ON 420 SUPPORT IMAGES

Class name	Precision	Recall	F1-score	Support
agricultural	1.00	1.00	1.00	20
airplane	1.00	0.85	0.92	20
baseballdiamond	0.89	0.80	0.84	20
Beach	1.00	0.95	0.97	20
buildings	0.64	0.70	0.67	20
chaparral	1.00	1.00	1.00	20
denseresidential	0.50	0.55	0.52	20
forest	0.86	0.95	0.90	20
freeway	1.00	0.90	0.95	20
golfcourse	0.80	0.60	0.69	20
harbor	0.87	1.00	0.93	20
intersection	0.88	0.70	0.78	20
mediumresidential	0.53	0.85	0.65	20
mobilehomepark	0.93	0.65	0.76	20
overpass	0.94	0.75	0.83	20
parkinglot	1.00	0.90	0.95	20
River	0.76	0.95	0.84	20
runway	0.87	1.00	0.93	20
sparseresidential	0.83	0.95	0.88	20
storagetanks	0.95	0.95	0.95	20
tenniscourt	0.94	0.80	0.86	20
Average performance	**0.87**	**0.85**	**0.86**	**420**

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3215264

TABLE III
TL CLASSIFICATION PERFORMANCE IN PRECISION, RECALL, AND F1-SCORE ON 420 SUPPORT IMAGES

Class name	Precision	Recall	F1-score	Support
agricultural	1.00	0.95	0.97	20
airplane	0.95	1.00	0.98	20
baseballdiamond	1.00	0.90	0.95	20
beach	0.95	0.95	0.95	20
buildings	0.90	0.90	0.90	20
chaparral	1.00	1.00	1.00	20
denseresidential	0.56	0.50	0.53	20
forest	0.82	0.90	0.86	20
freeway	1.00	0.95	0.97	20
golfcourse	0.92	0.55	0.69	20
harbor	0.91	1.00	0.95	20
intersection	0.83	0.95	0.88	20
mediumresidential	0.57	0.80	0.67	20
mobilehomepark	0.67	0.50	0.57	20
overpass	0.95	0.90	0.92	20
parkinglot	1.00	1.00	1.00	20
river	0.69	0.90	0.78	20
runway	1.00	0.95	0.97	20
sparseresidential	0.83	0.95	0.88	20
storagetanks	1.00	1.00	1.00	20
tenniscourt	1.00	0.80	0.89	20
Average performance	**0.88**	**0.87**	**0.88**	**420**

TABLE IV
FINE-TUNING CLASSIFICATION PERFORMANCE IN PRECISION, RECALL, AND F1-SCORE ON 420 SUPPORT IMAGES

Class name	Precision	Recall	F1-score	Support
agricultural	1.00	1.00	1.00	20
airplane	1.00	1.00	1.00	20
baseballdiamond	1.00	0.90	0.95	20
Beach	0.95	0.95	0.95	20
buildings	0.80	0.80	0.80	20
chaparral	1.00	1.00	1.00	20
denseresidential	0.62	0.65	0.63	20
Forest	0.86	0.95	0.90	20
freeway	1.00	1.00	1.00	20
golfcourse	0.80	0.60	0.69	20
Harbor	0.95	1.00	0.98	20
intersection	1.00	0.65	0.79	20
mediumresidential	0.66	0.95	0.78	20
mobilehomepark	0.61	0.55	0.58	20
overpass	0.91	1.00	0.95	20
parkinglot	0.95	1.00	0.98	20
River	0.73	0.95	0.83	20
Runway	1.00	1.00	1.00	20
sparseresidential	0.89	0.80	0.84	20
storagetanks	0.95	0.95	0.95	20
tenniscourt	1.00	0.80	0.89	20
Average performance	**0.89**	**0.88**	**0.89**	**420**

The accuracy of the DL models is also measured in terms of accuracy and loss measurement metrics in graphical representation, as shown in Figure 2, Figure 3, and Figure 4 for CNN-FE, TL, and fine-tuning models, respectively. The training accuracies (with a blue color curve) are smoothly increasing, while the validation accuracies (with a red color curve) are somewhat fluctuating in increasing the accuracies in all models, especially in fine-tuning, as depicted in Figure 2, Figure 3, and Figure 4 (on the **left**). We used the cross-entropy loss function to reduce errors in the model performance. The training losses (with a blue color curve) are smoothly decreasing, while the validation losses (with a red color curve) are somewhat fluctuating in reducing the errors in all models, as depicted in Figure 2, Figure 3, and Figure 4 (on the **right**).

In addition to deploying precision, recalls, and f1-score, we used CM to evaluate class performances in each DL model. Like f1-score, better class performance is observed in most classes in the CM metric. The CM measures the class performance, whether it is classified correctly or incorrectly. CM considers each class label in rows (True labeled class) and columns (predicted labeled class), as depicted in Figure 5 through Figure 7. The probability score in the diagonal intersection showed the correct classified classes. In contrast, the results in other rows-columns wise are predicted in misclassified classes.

For evaluating the model with test set sample images, we used the argmax function for predicting a class with the maximum argument probability score. For instance, our classification problem has twenty-one possible classes in the UCM dataset. If the output probabilities are [0.0, 0.0, 0.0, 0.0, 0.05, 0.0, 0.55, 0.0, 0.0, 0.0, 0.05, 0.0, 0.30, 0.05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0], the arg max (maximum argumentative-class) probability is 0.55 and it is associated with the denseresidential class that is predicted by the CNN-FE model as shown in Figure 5. Like ways, the arg max probability can correspond to each class prediction in the CM metric. The sum of the output probabilities of each class is **1.00.**

While evaluating the class performance in CM metric, the lowest performance results in the two first classes are denseresidential (55%) and golfcourse (60%) in CNN-FE, denseresidential (50%), mobilehomepark(50%) and golfcourse (55%) in TL and mobilehomepark(55%) and golfcourse (60%), as shown in Figure 5, Figure 6, and Figure 7, respectively. The lower result showed that the class property is mainly associated with other classes. For instance, the property of denseresidential has a more common feature with class mediumresidential. The class performance in the CM metric is generally better in the fine-tuning model.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3215264

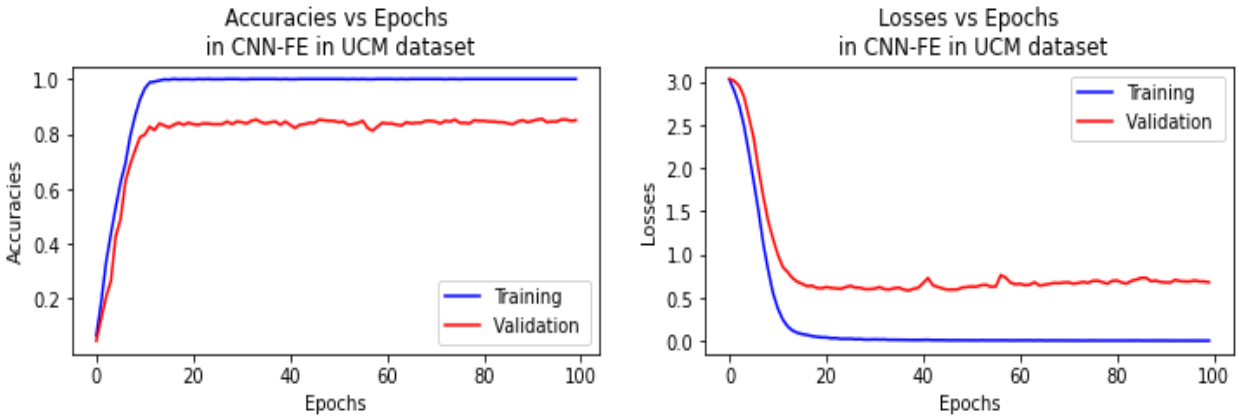


FIGURE 2. **The training and validation accuracies and losses in the CNN-FE model**

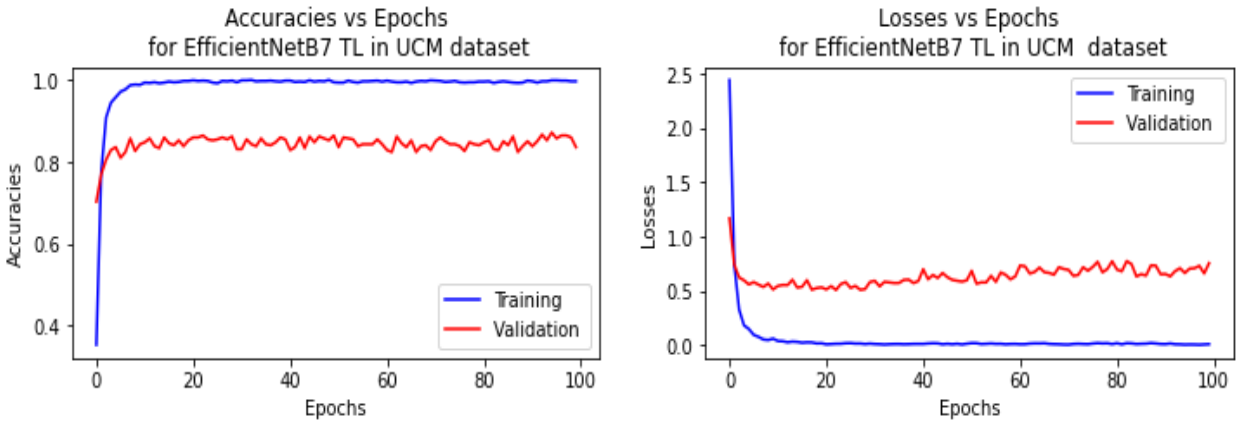


FIGURE 3. **The training and validation accuracies and losses in the TL model**

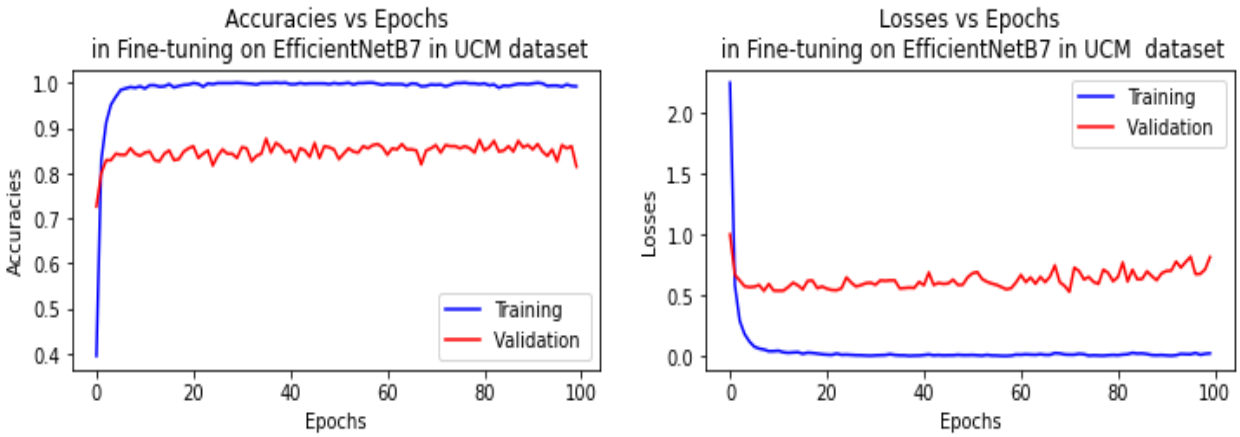


FIGURE 4. **The training and validation accuracies and losses in the fine-tuning DL model**

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3215264

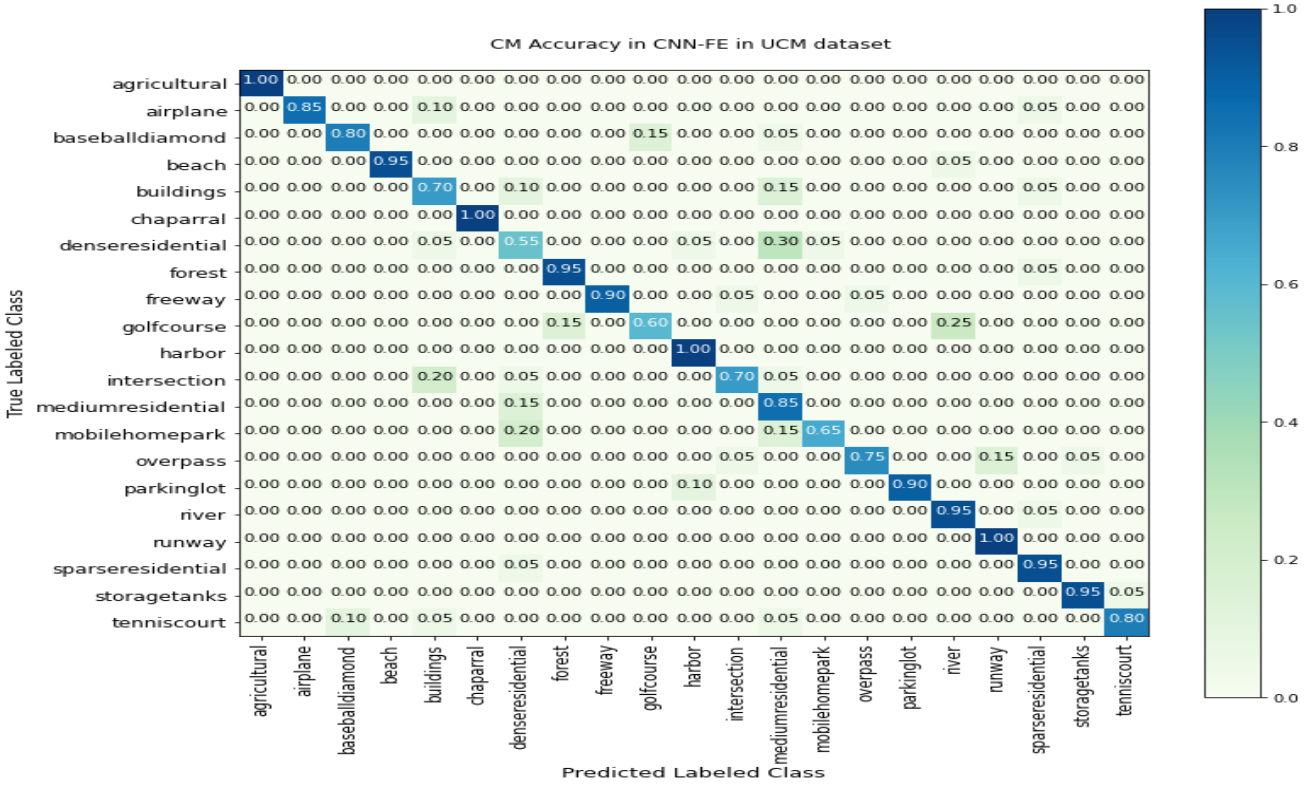


FIGURE 5. **CM performance results for the CNN-FE model in the UCM dataset**

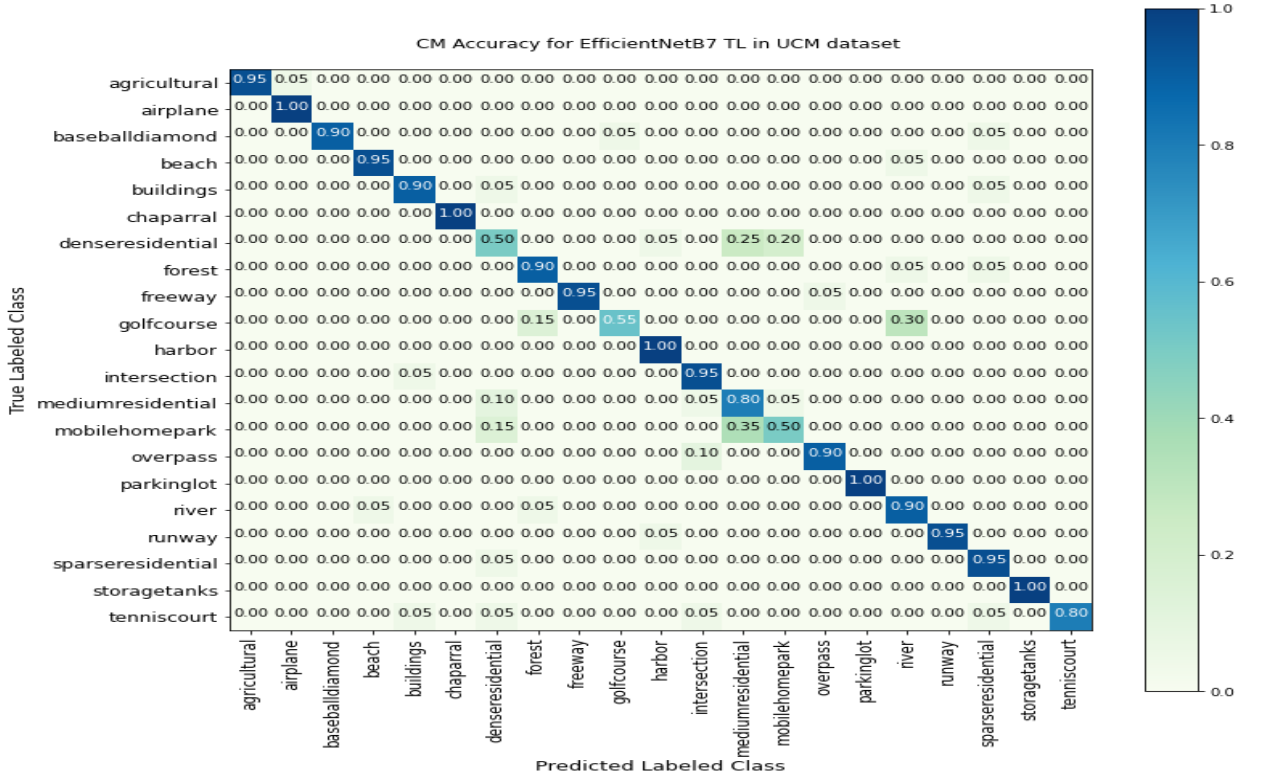


FIGURE 6. **CM performance results for the TL model in the UCM dataset**

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3215264

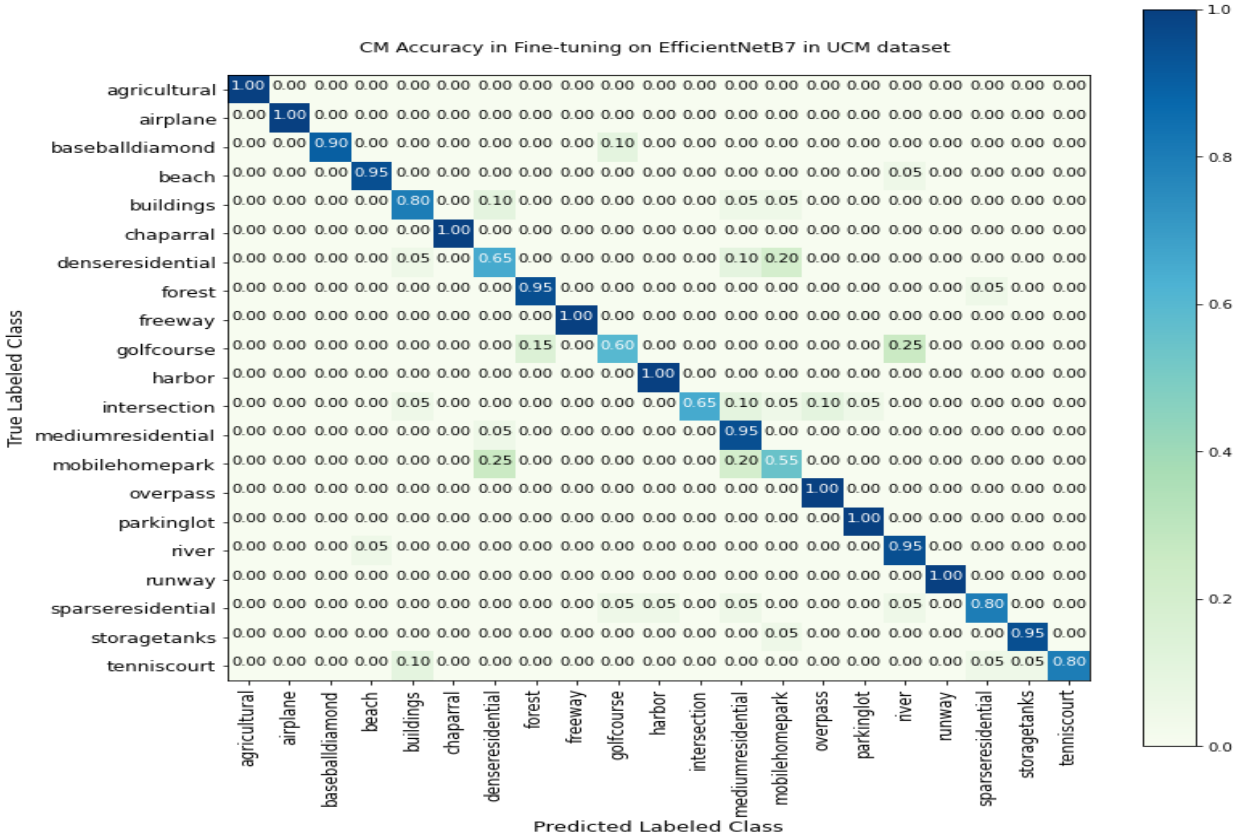


FIGURE 7. CM performance results for the fine-tuning model in the UCM dataset

B. DISCUSSIONS

This paper applied the DL models for LCLU classification using RS images. The performances of these models have been evaluated using various measurement metrics. The result showed good performance results for the classification problem, as shown in Table V. The experimental results showed that the proposed DL algorithms could adapt and learn features of remote sensing images since the Adam (adaptive movement estimation) learning rate takes the responsibility. The TL and fine-tuning performances are significantly improved from the scratch-developed CNN-FE.

To address our objective stated in this paper, Table II through Table IV and Figure 2 through Figure 7 compare the DL model performances on the UCM dataset. From the results, good class performance has been achieved in precision, recall, F1-score, and CM though some class performances scored lower in values. The training accuracy increases smoothly in CNN-FE, TL, and fine-tuning DL models, as shown in Figure 2, Figure 3, and Figure 4, respectively.

The overall accuracies for each model are summarized in Table V. According to Table V; the fine-tuning model has outperformed performance in accuracy (88%), precision (89%), recall (88%), and f1-score (89%) with efficient time. Whereas the CNN-FE model performed lower performances in each metric compared to the other two models, this could be why the dataset used was smaller. Moreover, the CNN-FE spent much more time training the model than the two.

The maximum capability of a number of parameters in EfficientNet is 64 million. There are 18.88 million parameters found and learned in the same number in TL and fine-tuning models. This parameter number is about 18 times greater than the parameters found in CNN-FE mode, i.e., 1.68 million. This is why the convolution technique used in CNN-FE reduces the number of parameters.

The fine-tuning technique is used to compare the performance of the DL models designed using the convolutional method and fully connected layers. As a result, improved performance has been achieved in fine-tuning model with efficient time compared to the other DL techniques used in this paper.

Designing the CNN model from scratch is essential to identify the correct properties of the categories for large datasets that are usually recommended when it exceeds about 5000 images per class. However, it could consume much training time and be subject to over-fitting. This limitation could be overcome by the less training time-consuming DL techniques, TL, and fine-tuning. The TL and fine-tuning DL techniques are efficient in training time consuming and produce improved performance results. But, we recommend that TL and fine-tuning DL be applicable for small data sizes that may be less than 5,000 images per class. Therefore, as observed in Table V, we can conclude that the TL and fine-turning DL techniques are economically relevant in time saving and essential for performance improvement.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3215264

TABLE V
THE DL MODEL PERFORMANCE EVALUATIONS USING PERFORMANCE MEASUREMENT METRICS IN THE UCM DATASET AND THE TIME (IN SECONDS) CONSUMED FOR TRAINING EACH DL MODEL

DL Models	DL Performance results in each measurement metrics				Training time (**Sec**.)	Params# (**millions**)
	Precision	Recall	F1-score	Accuracy		
CNN-FE	87.00	85.00	86.00	84.76	51.81	1.68
TL	88.00	87.00	88.00	87.38	41.77	88.88
Fine-tuning	**89.00**	**88.00**	**89.00**	**88.10**	**36.39**	18.88

IV. CONCLUSION

This paper designed the three DL models: CNN-FE, TL, and fine-tuning for LCLU classification problems using RS images. The TL and fine-tuning models have been trained on the recent EfficientNetB7 pre-trained baseline network, whereas the CNN-FE has been trained from scratch using the UCM dataset. The models' performances were evaluated using accuracy, precision, recall, f1-score, and CM metrics. The fine-tuned model in the UCM dataset has a profound accuracy result. We could observe that the nature of the denseresidential class is mainly similar to the properties of the mediumresidential category. Thus, its performance results in precision, recall, f1-score, and CM is the worst result compared to other class categories. In addition to those metrics, the training time is another critical evaluation metric that is used to compare the economic advantages of TL and fine-tuning models over the DL models developed from scratch. We found that the TL and fine-turning DL models are efficient in saving time and essential for performance improvements.

Since TL and fine-tuning modes are reliable for a small dataset, all the experiments have been performed using a small LCLU dataset in Google CoLab. Thus, we would like to recommend developing a DL model from scratch and comparing its performance with other pre-trained DL models using the powerful GPU to improve the performance of the DL models using other larger RS datasets. Like datasets, varying the DL hyperparameters could also affect the model performance. Thus, researching by considering the DL optimization techniques for LCLU classification using RS images would be our further study.

REFERENCES

[1] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. Benediktsson, "Deep Learning for Hyperspectral Image Classification : An Overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, 2019.

[2] B. Liu, X. Yu, P. Zhang, A. Yu, Q. Fu, and X. Wei, "Supervised Deep Feature Extraction for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 1909–1921, 2018.

[3] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring Hierarchical Convolutional Features for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6712–6722, 2018.

[4] M. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanesh, and Y. Zhang, "Very Deep Convolutional Neural Networks for Complex Land Cover Mapping using Multispectral Remote Sensing Imagery," *Remote Sens.*, vol. 10, no. 7, p. 1119, 2018.

[5] M. A. Shafaey, M. A. Salem, H. M. Ebied, M. N. Al-Berry, and M. F. Tolba, "Deep Learning for Satellite Image Classification," *Springer Nat. Switz.*, vol. AISI 2018, no. AISC 845, pp. 383–391, 2019.

[6] G. J. Scott, M. R. England, W. A. Starms, R. A. Marcum, and C. H. Davis, "Training Deep Convolutional Neural Networks for Land-Cover Classification of High-Resolution Imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 549–553, 2017.

[7] A. Hamida, A. Benoit, P. Lambert, and C. Amar, "3-D Deep Learning Approach for Remote Sensing Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4420–4434, 2018.

[8] C. Deng, Y. Xue, X. Liu, C. Li, and D. Tao, "Active Transfer Learning Network: A Unified Deep Joint Spectral-Spatial Feature Learning Model for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1741–1754, 2019.

[9] F. Özyurt, "Efficient Deep Feature Selection for Remote Sensing Image Recognition with Fused Deep Learning Architectures," *J. Supercomput.*, vol. 76, no. 11, pp. 8413–8431, 2019.

[10] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, "Object Detection in Optical Remote Sensing Images based on Weakly Supervised Learning and High-Level Feature Learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3325–3337, 2015.

[11] O. Fink, Q. Wang, M. Svensén, P. Dersin, W.-J. Lee, and M. Ducoffe, "Potential, Challenges and Future Directions for Deep Learning in Prognostics and Health Management Applications," *Eng. Appl. Artif. Intell.*, vol. 92, no. May, p. 103678, 2020.

[12] X. Liu, Y. Zhou, J. Zhao, R. Yao, B. Liu, and Y. Zheng, "Siamese Convolutional Neural Networks for Remote Sensing Scene Classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1200–1204, 2019.

[13] U. Zahid *et al.*, "BrainNet : Optimal Deep Learning Feature Fusion for Brain Tumor Classification," *Comput. Intell. Neurosci.*, vol. 2022, p. 1465173, 2022.

[14] B. Yang, S. Hu, Q. Guo, and D. Hong, "Multisource Domain Transfer Learning Based on Spectral Projections for Hyperspectral Image Classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 15, pp. 3730–3739, 2022.

[15] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, 2017.

[16] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *arXiv Prepr. arXiv1905.11946v1*, pp. 1–10, 2019.

[17] A. Shabbir *et al.*, "Satellite and Scene Image Classification Based on Transfer Learning and Fine Tuning of ResNet50," *Math. Probl. Eng.*, vol. 2021, p. 5843816, 2021.

[18] H. Alhichri, A. S. Alswayed, Y. Bazi, N. Ammour, and N. A. Alajlan, "Classification of Remote Sensing Images Using EfficientNet-B3 CNN Model with Attention," *IEEE Access*, vol. 9, no. 2021, pp. 14078–14094, 2021.

[19] Y. Yang and S. Newsam, "Bag-of-Visual-Words and Spatial Extensions for Land-Use Classification," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010, pp. 270–279.

[20] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.

[21] R. P. de Lima and K. Marfurt, "Convolutional Neural Network for Remote - Sensing Scene Classification : Transfer Learning Analysis," *Remote Sens.*, vol. 12, no. 1, p. 86, 2020.

[22] R. Stivaktakis, G. Tsagkatakis, and P. Tsakalides, "Deep Learning for Multilabel Land Cover Scene Categorization Using Data

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3215264

Augmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1031–1035, 2019.

[23] C. Peng, Y. Li, L. Jiao, Y. Chen, and R. Shang, "Densely Based Multi-Scale and Multi-Modal Fully Convolutional Networks for High-Resolution Remote-Sensing Image Semantic Segmentation," *IEEE J. Sel. Top. Appl. EARTH Obs. Remote Sens.*, vol. 12, no. 8, pp. 2612–2626, 2019.

[24] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Fully convolutional neural networks for remote sensing image classification," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 5071–5074.

[25] Z. Zeng, X. Chen, and Z. Song, "MGFN: A Multi-Granularity Fusion Convolutional Neural Network for Remote Sensing Scene Classification," *IEEE Access*, vol. 9, pp. 76038–76046, 2021.

[26] M. Kim *et al.*, "Convolutional Neural Network-Based Land Cover Classification Using 2-D Spectral Reflectance Curve Graphs With Multitemporal Satellite Imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 12, pp. 4604–4617, 2018.

[27] R. Naushad and T. Kaur, "Deep Transfer Learning for Land Use Land Cover Classification : A Comparative Study," *arXiv:2110.02580v2, Online available at https://arxiv.org/abs/2110.02580*, 2021.

[28] D. Zhang, Z. Liu, and X. Shi, "Transfer learning on EfficientNet for remote sensing image classification," in *Proceedings - 2020 5th International Conference on Mechanical, Control and Computer Engineering, ICMCCE 2020*, 2020, pp. 2255–2258.

[29] A. Alem and S. Kumar, "Transfer Learning Models for Land Cover and Land Use Classification in Remote Sensing Image," *Appl. Artif. Intell.*, vol. 60, no. 1, p. 2014192, 2022.

[30] M. Rashid *et al.*, "A Sustainable Deep Learning Framework for Object Recognition using Multi-layers Deep Features Fusion and Selection," *Sustain.*, vol. 12, no. 12, p. 5037, 2020.



ABEBAW ALEM received the BSc. and MSc. degrees in information science from Haramaya University, Haramaya, Oromia, Ethiopia, in 2010, and from Addis Ababa University, Addis Ababa, Ethiopia, in 2014, respectively. He is currently pursuing his Ph.D. degree in computer science and engineering at Delhi Technological University, India.

From 15/9/2010 to 18/7/2018, he was a faculty member of the Computer Science and IT Department, Debre Tabor University, Debre Tabor, Ethiopia. He also has middle-level management experience as Vice Dean of Technology Faculty from 06/11/2014 to 30/10/2016 and Technology Faculty Educational Development Center Coordinator from 31/10/2016 to 18/7/2018. His research interest is Deep Learning for Remote Sensed Image Classification, Machine Learning for Image Analysis, Case-Based Reasoning, and Computer Vision.

Mr. Alem is a member of the Black Artificial Intelligence professional group, Global Initiative of Academic Networks (GIAN), and Young African Leaders Initiative (YALI).



SHAILENDER KUMAR received a Bachelor of Engineering (Computer Science and Engineering) in 2001, a Master of Technology (Computer Science) in 2005, and a Doctorate of Philosophy in Computer Science and Engineering from Maharshi Dayanand University, Haryana, India, in 2017.

He has more than 20 years of teaching experience in the computer science and engineering department at various esteemed Engineering Colleges like Delhi College of Engineering, Netaji Subhas Institute of Technology, and Ambedkar Institute of Advanced Communication Technologies and Research, Delhi, India. Moreover, he has positional experiences, such as being an Officer-in-Charge of Data Mining Lab and Ph.D. students coordinator in the computer science and engineering department at Delhi Technological University, Delhi, India. His research interest is Database Systems, Data mining, Big Data Analytics, Machine Learning, Information Security, Compiler Construction, and Computer Network.

Prof. (Dr.) Kumar is currently a professor of Computer Science and Engineering at Delhi Technological University, Delhi, India. He has published over 50 research papers in various reputed International Journals and Conferences.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

2022 2nd Asian Conference on Innovation in Technology (ASIANCON)
Pune, India. Aug 26-28, 2022

Design and Analysis of LLC Resonant Converter and CCCV Topology for Battery Charging

Praphull Prakash Gupta
Department of Electrical Engineering, Delhi Technological University
New Delhi,India
praphullprakashgupta_2k20pes14@dtu.ac.in

Narendra Kumar
Department of Electrical Engineering, Delhi Technological University
New Delhi,India
narendrakumar@dtu.ac.in

Uma Nangia
Department of Electrical Engineering, Delhi Technological University
New Delhi,India
umanangia@dce.ac.in

2022 2nd Asian Conference on Innovation in Technology (ASIANCON) | 978-1-6654-6851-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ASIANCON55314.2022.9909135

***Abstract*— The LLC resonant converters are widely used in various sectors of the electronics-based industries due to their advantages of high efficiency, high energy density, electrical isolation, low electromagnetic interference (EMI),wide output ranges and high operation frequency. A Good power factor for the LLC resonant tank is required to achieve high efficiency over a wide input voltage range. A unity voltage gain can be accomplished over the whole loading conditions by having switching frequency equal to resonant tank frequency in LLC converter, however it is not suited for constant current – constant voltage (CC-CV) battery applications. The use of a buck converter at the output of an LLC resonant converter to regulate the charging current and voltage for battery charging applications is proposed in this paper. The use of a buck converter allows the output voltage of the LLC converter to be varied, ensuring CCCV charging. The proposed system is analyzed and studied in MATLAB/Simulink.**

Keywords—LLC converter, Resonant tank, Buck converter, zero voltage switching(ZVS), CCCV topology, close loop control

I. INTRODUCTION

Battery charger circuits are required by the widespread use of batteries in all area of power systems, such as cell phones and hybrid electric vehicles (EVs), and renewable energy systems (RESs). The main requirements for these circuits are that the power electronic unit has a high-power density and operates at high efficiencies, as well as the ability to keep up with the battery pack's dynamic behaviour[1].

Reducing the bulk of switching power supply by increasing the switching frequency has become a global trend in recent years. However, substantial switching losses result in low efficiency due to the high frequency. The resonant converter has been widely employed in the power industry because it has a ZVS or ZCS function for decreasing switching losses. The resonant LLC DC-DC converter[2].

The Resonant tank is series combination of resonant inductor and resonant capacitor parallel combination of magnetizing inductance and equivalent resistance The LLC resonant tank's voltage gain curves are critical. The Switch attains the zero-voltage switching (ZVS) in turn-on condition. when the operating frequency is higher than the frequency at peak voltage gain of LLC resonant tank with diverse load conditions[2].

In LLC converters, a sophisticated control technique, such as variable frequency control, is frequently used. Other advanced control systems for wide-range input or output voltage applications are also possible. The input voltage was set to 400 V for this paper application. There is no need to consider a large range of input voltage when designing the LLC converter. In such scenario, the LLC converter with open loop control is cascaded with a buck converter as its output to regulate the charging current and voltage in CC-CV battery charging. The CC-CV battery charger will be simple to implement[3].

Constant voltage, constant current and a combination of constant voltage and constant current with or without a smart charging circuit are the three most prevalent methods of charging a battery.

Constant current charging is more efficient than constant voltage charging since the charging current in constant voltage charging might be high, causing the battery to overheat during charging. Voltage charging, on the other hand, is faster than constant current charging. The two sorts of sources are used in the practical charging procedure. When the battery is relatively depleted, continuous current charging is used. When the battery reaches a specific voltage, it is close to its maximum voltage. It is possible to charge at a constant voltage[4].

As a result, the LLC converter's output voltage may not be load independent. The use of a buck converter to regulate the output of an LLC converter allows the output voltage of the LLC converter to vary, ensuring CC-CV charging.

A. Modelling of LLC Converter

On-board, a resonant converter with a resonant tank is commonly used to provide isolation for safety and a high output power level[2].The entire bridge is one of the most well-known topologies.

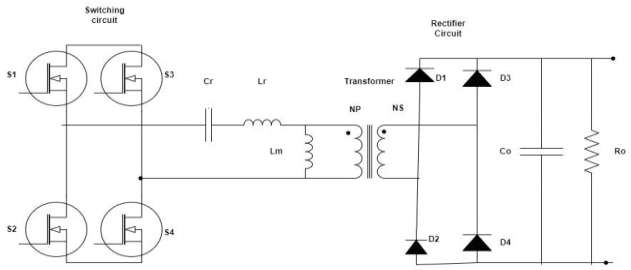


Fig. 1. circuit diagram of Full bridge LLC converter with Full bridge rectifier

By analyzing the above circuit (Figure1), the equivalent circuit will be obtained.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:12:49 UTC from IEEE Xplore. Restrictions apply.

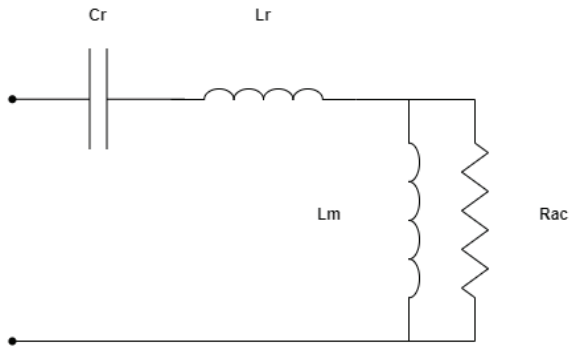


Fig. 2. Equivalent circuit of LLC resonant tank

The following formulas can be used to calculate the magnitude of the analogous circuit's output – input voltage ratio[5].

$$K(Q,m,F_1)=\left|\frac{V_{o_ac}(s)}{V_{in_ac}(s)}\right|=\frac{F_x^{\,2}(m-1)}{\sqrt{(m.F_x^{\,2}-1)^2+F_x^{\,2}.\left(F_x^{\,2}-1\right)^2.(m-1)^2.Q^2}} \quad (1)$$

$R_{ac}=\frac{8N_P^2}{\pi^2N_S^2}.R_o$ Reflected load resistance

$Q=\frac{\sqrt{L_r/C_r}}{R}$ Q is Quality factor

$F_N=\frac{f_s}{f_r}$ Normalized Frequency

$f_r=\frac{1}{2\pi\sqrt{L_r.C_r}}$ Resonant frequency

$m=\frac{L_r+L_m}{L_r}$ Ratio of total primary inductance to resonant inductance.

TABLE I. CIRCUIT PARAMETERS OF LLC CONVERTER

Parameters	**Values**
Input voltage	400 V
Rectified Output voltage of LLC	95 V
Rectified output current of LLC	8.3 A
Resonant Inductance Lr	5.16 mH
Resonant Capacitance Cr	491.39 PF
Resonant Frequency Fr	100 KHz
Switching frequency Fs	100 KHz
Magnetizing Inductance Lm	25.86 mH
Turn ratio (Np:Ns)	4:1
R_o	12 Ω

The R_o value for this application is going to supply a lithium–ion battery with 48 V/40 Ah capacity. The constant current for this to be 4 Ampere and the minimum voltage for the battery is set to 50 V. close loop control gives the output of 49.6 V constant voltage and 4.12 A constant current which suitable to charge 48 V/40Ah battery.

II. PROPOSED TOPOLOGY

The voltage gain stays 1 regardless of load situation since the switching frequency is the same as the resonant frequency. The LLC converter therefore acts as a fixed output voltage source for the buck converter. The duty cycle of the buck converter for the CC- CV charging Battery is varied[3].

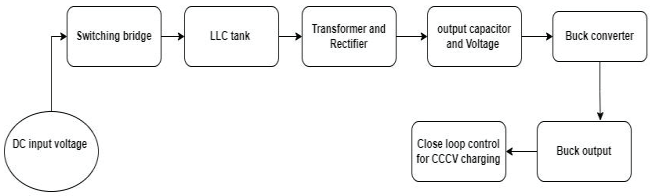


Fig. 3. Block diagram of LLC converter with Buck converter

The Block diagram depicts the DC voltage given to the switching bridge which is bridge circuit. The LLC Resonant circuit gives the Zero voltage switching (ZVS) in the primary side. ZVS reduces the switching losses in Proposed circuit. LLC output is given to the Buck converter, by this we can regulates the charging current and charging voltage for our battery specification. We will regulate the voltage and current by varying the duty cycle of Buck Converter. Close loop control of current and voltage will give the Constant current (CC) and Constant voltage (CV).

TABLE II. CIRCUIT PARAMETERS OF BUCK CONVERTER

Parameters	**Values**
Inductance	30 uH
Capacitance	30.4 uF
Duty ratio	0.5
Switching Frequency	100 KHz
Load resistance	12 Ω

Another advantage of employing a Buck converter is that the demand for LLC converter transformer design is removed. As a result, the transformer's main and secondary windings might be wound on two different cores. It allows for the creation of a contactless connector for an electric boat charger. Because people are concerned about getting an electric shock near the water when charging an electric boat with a metal conductive charger connector, it would be beneficial if a contactless charger connector could be developed[6].

A. Battery Charging using CCCV Topology

As illustrated in Fig. 4, the battery charging voltage is identical to the battery actual voltage in CC mode. As a result, in this mode, the battery should experience a voltage drop across it that is equivalent to its real voltage. When the battery voltage hits 4.0 V, the battery must be supplied a constant voltage of 4.2 V, which is the battery's maximum rated voltage. When the battery's SOC level hits 100% or the voltage reaches 4.2 V, the charging current should start to drop, and the battery should be regarded completely charged.[7].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:12:49 UTC from IEEE Xplore. Restrictions apply.

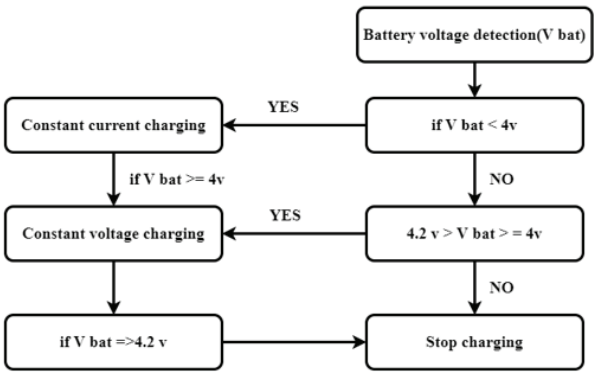


Fig.4: Flow chart of charging algorithm

We'll monitor the battery cell voltage and charge it with continuous current if it's less than 4V (less than 95 percent SOC). We'll switch the charging mode to constant voltage after the battery cell voltage reaches 4 V to charge the battery cell from 4V to 4.2 V (maximum cell voltage)[8]

To avoid overcharging, we will disconnect the switch when the battery cell SOC hits 100% or the battery cell voltage reaches 4.2 V.

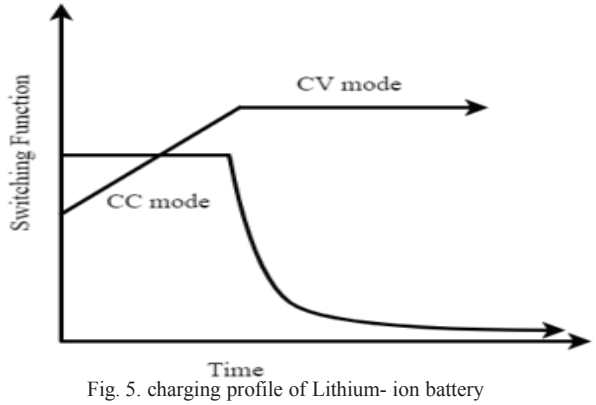


Fig. 5. charging profile of Lithium- ion battery

A CCCV charging cycle for a single lithium-ion cell is depicted in Fig. 5. We change the function from constant current to constant voltage once SOC reaches 95%. In CV mode, very little current flows, and from 95 percent to 100 percent SOC, current drops to 0 ampere[9][10].

B. close loop control of Current and voltage

When the resonant frequency equals the switching frequency (fs=fr), the LLC converter has a constant characteristic. The output voltage of the LLC converter is seen to be independent of load state. It is unavoidable to use a PID controller to control the switching mode of CC-CV mode charging when employing closed-loop control. PI compensation was used to implement the CC-CV mode charging operation based on voltage and current feedback. In close loop control voltage is control by PI controller and current is controlled by Hysteresis control[11].

TABLE III. OUTPUT VALUES OF BUCK CONVERTER WITH AND WITHOUT CLOSE LOOP CONTROL

Parameters	**Values**
Output voltage of Buck	48.7 V
Output current of Buck	4.06 A
Close loop voltage	49.6 V
Close loop current	4.13 A

III. SIMULATION RESULTS

A. Zero voltage switching of LLC converter

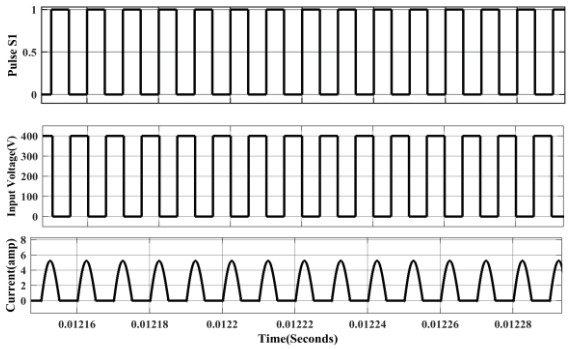


Fig. 6: ZVS of LLC converter

When switch S1 is on, at this time voltage across the switch becomes zero and current start from the negative value. This is perfect ZVS condition[12][13]. ZVS will reduces the switching losses of the converter and increases the efficiency [14].

B. Secondary side voltage of Transformer

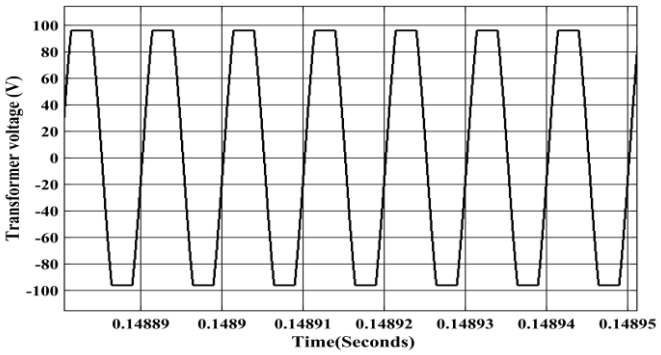
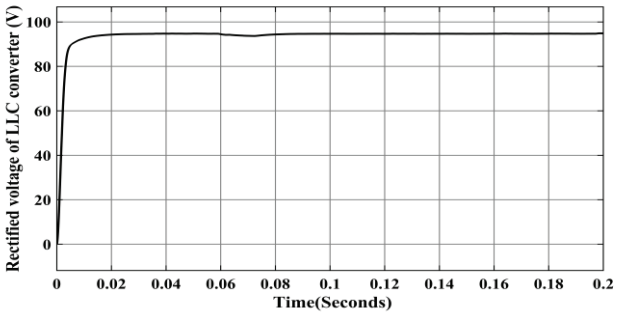


Fig. 7: Transformer output voltage

Figure 7 depicts the transformer's secondary voltage side. The transformer has a 4:1 turn ratio. It's kind of similar to a square wave.

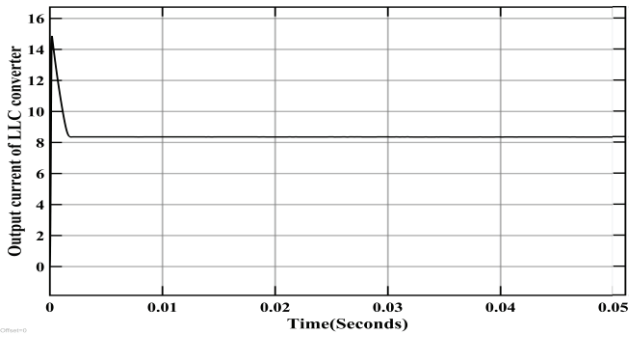
The transformer output voltage is given to the full bridge rectifier which will give the rectified volage and current.

C. Rectified Output Voltage and current LLC converter



(a)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:12:49 UTC from IEEE Xplore. Restrictions apply.

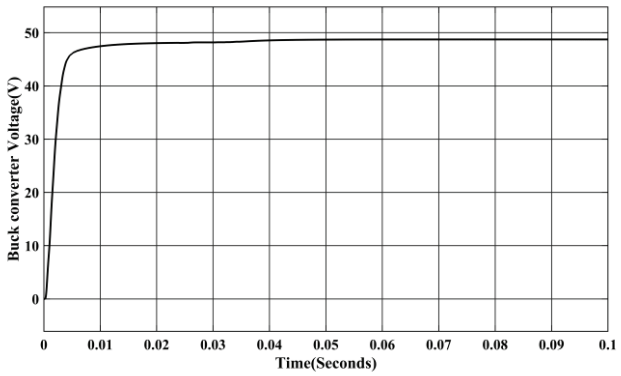


(b)

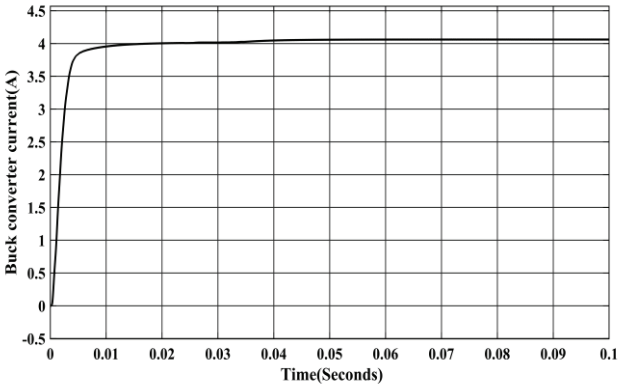
Fig. 8: (a) Rectified Output voltage of LLC converter (b) Rectified Output current of LLC converter

Figure 8 (a) depicts the Rectified voltage of LLC converter with Buck converter in cascade and current is around 8.33 A as seen in figure 8 (b),Voltage is less than 100 V because of non -linear components of Buck converter.

D. Output current and voltage of Buck converter



(a)

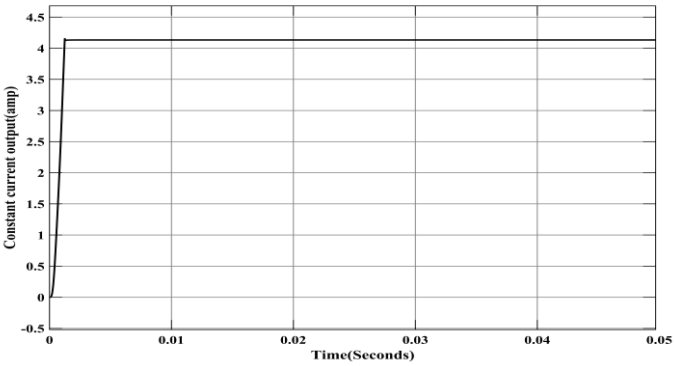


(b)

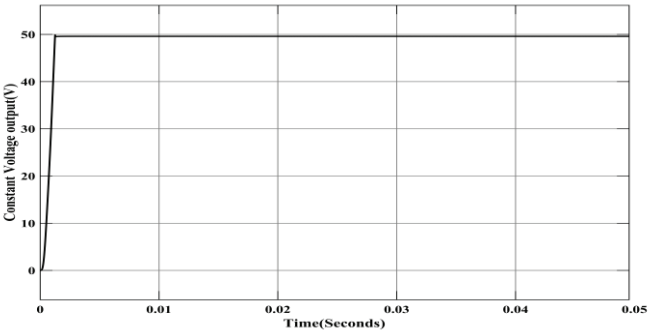
Fig. 9:(a) Output current of Buck converter (b) Output Voltage of Buck converter

Buck converter is giving 48.7 V regulated constant voltage by using a duty cycle of 0.5 as shown in figure 9 (a) and Buck converter output current is 4.06 A as per figure 9 (b). for constant current – constant voltage charging we have to do close loop control of Buck converter output voltage and current as per our battery specifications.

E. Close loop control for CC- CV charging



(a)



(b)

Fig. 10: (a) closed loop constant current (b) closed loop constant voltage

Buck converter output current is controlled by close loop of current by hysteresis current control and Buck converter output voltage is controlled by using PI controller. It is dual loop combination of PI controller and Hysteresis current control. Close loop control current is coming 4.13 A constant in figure 10 (a) and close loop voltage is coming constant of 49.6 V as shown in figure 10 (b).

IV. CONCLUSION

This paper presents an LLC converter with Buck converter for CC- CV charging. LLC converter improves the switching losses by Zero voltage switching. The simulation is presented for proving that the idea is Achievable and Suitable for CC-CV charging for a battery with Buck converter of efficiency 95 %. The Buck converter gives the option to get required output current and voltage as per the battery specifications. Close loop current and voltage is coming 4.13 A and 49.6 V which is perfectly suitable for charging a 48 V /40 Ah Battery. The closed loop control of current and voltage gives the Constant current and constant Voltage (CC-CV) for charging the Battery as per Battery specifications.

REFERENCES

[1] T. N. Gücin, M. Biberoğlu, and B. Fincan, "Constant frequency operation of parallel resonant converter for constant-current constant-voltage battery charger applications," *J. Mod. Power Syst. Clean Energy*, vol. 7, no. 1, pp. 186–199, 2019, doi: 10.1007/s40565-018-0403-7.

[2] R. L. Lin and C. W. Lin, "Design criteria for resonant tank of LLC DC-DC resonant converter," *IECON Proc. (Industrial Electron. Conf.*, no. 1, pp. 427–432,2010, doi: 10.1109/IECON.2010.5674988.

[3] H. K. T. Tsang, Y. C. Fong, K. L. J. Kan, S. R. Raman and K. W. E. Cheng, "A Study of LLC Converter with Buck Converter for CC-CV Charging," 2020 8th International Conference on Power

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:12:49 UTC from IEEE Xplore. Restrictions apply.

Electronics Systems and Applications (PESA),2020,pp.1-,doi:10.1109/PESA50370.2020.9

[4] J. Zeng, G. Zhang, S. S. Yu, B. Zhang, and Y. Zhang, "LLC resonant converter topologies and industrial applications -A review," *Chinese J. Electr. Eng.*, vol. 6, no. 3, pp. 73–84, 2020, doi: 10.23919/CJEE.2020.000021.

[5] V. B. Vu, D. H. Tran, and W. Choi, "Implementation of the Constant Current and Constant Voltage Charge of Inductive Power Transfer Systems with the Double-Sided LCC Compensation Topology for Electric Vehicle Battery Charge Applications," *IEEE Trans. Power Electron.*, vol. 33, no. 9, pp. 7398–7410, 2018, doi: 10.1109/TPEL.2017.2766605.

[6] S. Kivrak, T. Ozer, and Y. Oguz, "Battery Management System Implementation with Pasive Control Method," *2018 4th Int. Conf. Inf. Technol. Eng. Educ. Inforino 2018 - Proc.*, pp. 1–4, 2018, doi: 10.1109/INFORINO.2018.8581758.

[7] Q. Ouyang, Z. Wang, K. Liu, G. Xu, and Y. Li, "Optimal Charging Control for Lithium-Ion Battery Packs: A Distributed Average Tracking Approach," *IEEE Trans. Ind. Informatics*, vol. 16, no. 5, pp. 3430–3438, 2020, doi: 10.1109/TII.2019.2951060.

[8] Y. Shang, C. Zhu, Y. Fu, and C. C. Mi, "An Integrated Heater Equalizer for Lithium-Ion Batteries of Electric Vehicles," *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4398–4405, 2019, doi: 10.1109/TIE.2018.2863187.

[9] H. N. Vu and W. Choi, "A Novel Dual Full-Bridge LLC Resonant Converter for CC and CV Charges of Batteries for Electric Vehicles," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2212–2225, 2018, doi: 10.1109/TIE.2017.2739705.

[10] A. B. Khan and W. Choi, "Optimal Charge Pattern for the High-Performance Multistage Constant Current Charge Method for the Li-Ion Batteries," *IEEE Trans. Energy Convers.*, vol. 33, no. 3, pp. 1132–1140, 2018, doi: 10.1109/TEC.2018.2801381.

[11] L. Patnaik, A. V. J. S. Praneeth, and S. S. Williamson, "A Closed-Loop Constant-Temperature Constant-Voltage Charging Technique to Reduce Charge Time of Lithium-Ion Batteries," *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 1059–1067, 2019, doi: 10.1109/TIE.2018.2833038.

[12] A. S. Kumar and P. K. Reddy, "Design & Implementation of Zero Voltage Switching Buck Converter," *Ijera, Issn*, vol. 4, no. 9, pp. 193–198, 2014,[Online].Available:http://www.academia.edu/download/35316984/AG4905193198.pdf.

[13] M. M. Masoom, N. Kumar, and A. R. Kulkarni, "MPPT Based Grid Connected Photovoltaic System Using Flyback Converter," *2020 IEEE Int. Conf. Innov. Technol. INOCON 2020*, pp. 6–10, 2020, doi: 10.1109/INOCON50539.2020.9298412.

[14] Rashid, M. H., Kumar, N., & Kulkarni, A. (2014). Power Electronics Devices. *Circuits and Applications (4th Edition), England*.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:12:49 UTC from IEEE Xplore. Restrictions apply.

Design and Performance Analysis of High-Performance Low Power Voltage Mode Sense Amplifier for Static RAM

Divya DUTT [1], *Poornima MITTAL* [1], *Bhawna RAWAT* [1], *Brijesh KUMAR* [2]

[1]Department of Electronics and Communication Engineering, Delhi Technological University, Main Bawana Road, 110042 New Delhi, Delhi, India
[2]Department of Information Technology, Indira Gandhi Delhi Technical University for Women, Madrasa Road, 110006 New Delhi, Delhi, India

divyaa.dutt13@gmail.com, poornimamittal@dtu.ac.in, bhawnarawat12@gmail.com, brijeshkr@igdtuw.ac.in

DOI: 10.15598/aeee.v20i3.4373

Article history: Received Oct 04, 2021; Revised Apr 29, 2022; Accepted May 06, 2022; Published Sep 30, 2022.
This is an open access article under the BY-CC license.

Abstract. *In the prominent era of the digital world and Very Large-Scale Integration (VLSI) circuits, Static Random Access Memory (SRAM) provides a vital contribution to low-power and high-speed performance. Sense Amplifiers (SA) are a part of Complementary Metal-Oxide-Semiconductor (CMOS) memories used to read the stored information. This paper indicates a Dual-Voltage, Dual-Tail Level Restoration Voltage Latch Sense Amplifier (DVDTLR-VLSA). The design has been implemented using the LT SPICE tool at* 180 *nm technology node with a* 1.8 *V supply. Performance comparison of existing SA presented in literature with the proposed SA is examined based on different parameters like power, energy, delay, and current. The proposed design maintains power at* 2.167 *µW that is decreased to half as against Dual Switch Transmission Gate Voltage SA (DTGVSA) and shows an appreciable depletion. Also, the current and delay results are improved. Dimensional analysis is also done for the proposed SA to examine the performance. After that, the effect of sleep transistors on the proposed SA examines the performance in comparison to delay and power parameters without sleep transistors. The DVDTLR-VLSA has minimal energy and power. Also, the delay is improved which may be determined more advisable for low-power operations.*

Keywords

Aspect ratio, average output current, bitline, energy, level restoration, sense amplifier, sleep transistor, transmission gate.

1. Introduction

In the raising interest of high-speed and low-power VLSI circuits, Sense Amplifier (SA) has drawn the attention for enormous usage in CMOS memory and circuits. The SRAM circuit may explosively influence the power, delay, and stability of the design. Thus, SRAM cache is among the critical components of the modern VLSI system, and SRAM circuit is a salient element for portable electronic devices [1]. A permanent memory retains its data when power is disconnected after activation which is embedded inside "System On Chip" (SOC) circuits to provide high-level integration. SRAMs are used in consumer appliances applications since they can deeply influence the system performance concerning energy consumption, speed, and reliability [2]. It means they may perform high-speed and low-power operations. However, the major problem of SRAM circuits is the engagement of equilibrium between the writing and reading. As suggested in the literature, the cell ratio should be significant enough to permit the read access aside from changing the data [3] and [4].

SA is an essential part of the peripheral circuit for SRAM as it identifies and reads the information stored inside the bit cell during the action as far as the reading operation is completed. SA senses the bit lines' low signal voltage differences and amplifies them to the desired level, but it functions only while data is to be read in the memory. SAs are employed to amplify the low signal difference to the preferable logic level [5]. The differential SAs also called Voltage Mode SA (VMSA) are used as the static current is not passed through them. Therefore, to raise the performance and speed of memory, it is mandatory to identify and analyse various VMSAs' importance and benefits. The cross-coupled latch SAs are the most used and known due to their power and speed performance. However, it must be deliberately controlled since the input line also serves as the output line [6]. Between those SAs, the proposed SA has minimum power.

In this research work, a modified dual-voltage dual-tail level restoration voltage latch SA is designed at a 180 nm technology node that reveals low energy, delay, and power performance in comparison with any cross-coupled latch SA reported in the literature. Section 2. describes a literature review of the existing VMSA topologies. Section 3. presents the schematic design, characteristics analysis, dimensional analysis, and effect of sleep transistors on the proposed SA. A comparative analysis of the performance of the proposed SA with the pre-existing SAs is presented in Sec. 4. The findings of the paper are concluded in Sec. 5.

2. Existing Voltage-Mode Based SA Topologies

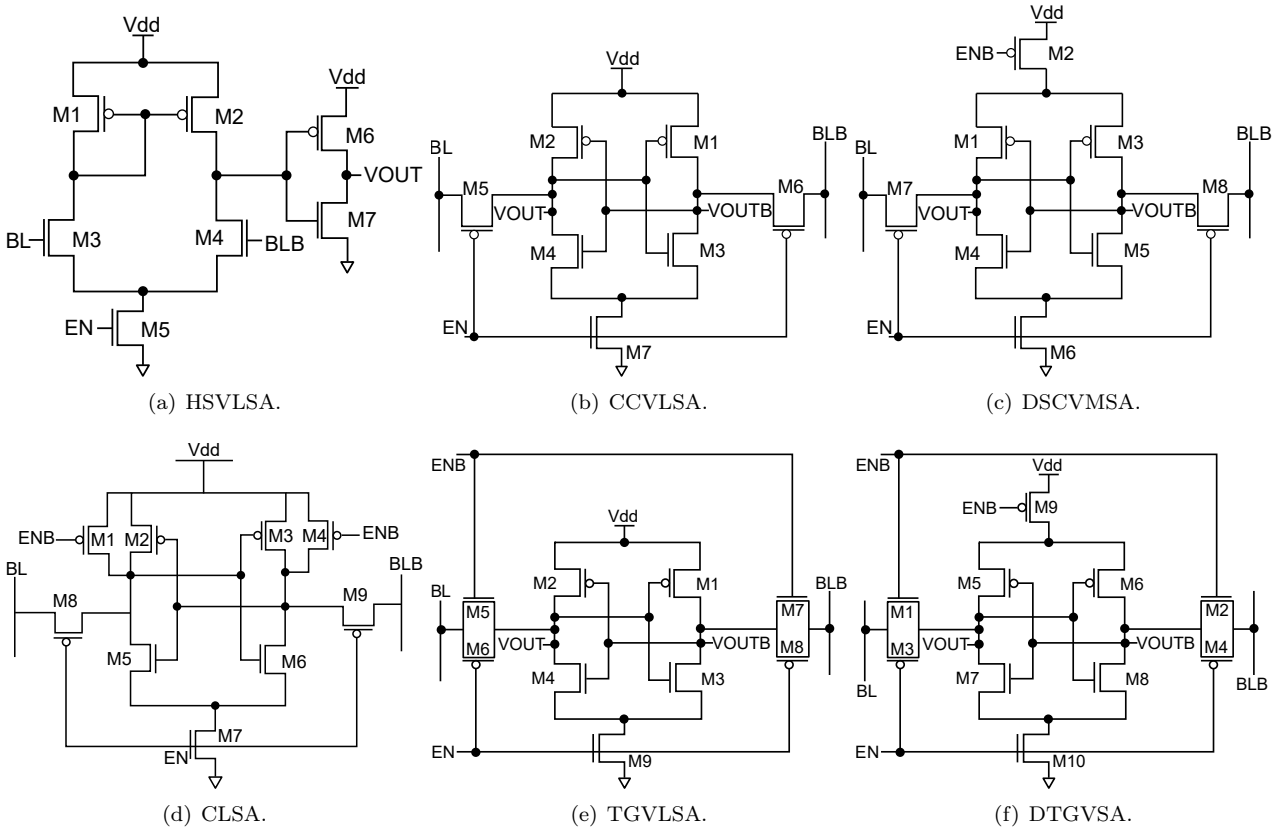
In this section, some well-known existing SA topologies reported in previous years are presented. The SA is a prerequisite ingredient that assembles the memory Integrated Circuit (IC) chip in the semiconductor industry. SA is the part of reading circuit systems used for data reading in the memory. Thus, it is a necessary component of functionality and performance. The SAs need reading data within the memory cells. The fundamental role of the SA is to sense the stored data based on bitline voltage discharge and to amplify the small voltage signal difference to a preferable voltage level. Thus, the stored data can be demonstrated correctly by logic aside from the memory [7]. The familiar and superior SA is SRAM based on cross-coupled voltage latch SA. Few existing SAs that achieve the desired ambition are shown in Fig. 1. The correlative property discussed in the paper for each SA is based on latch type of cross-coupled voltage mode design. W/L (Width/Length) ratio of each topology is equal i.e., 4 : 1 for all Positive channel Metal-Oxide-Semiconductor (PMOS) and 2 : 1 for NMOS transistors.

The basic High-Speed Voltage Latch SA (HSVLSA) design was reported in 2016 [8] and is shown in Fig. 1(a). In this circuit, there are three Negative channel Metal-Oxide-Semiconductor (NMOS) transistors (M3, M4, and M6) and two PMOS transistors (M1 and M2), and a CMOS inverter. The operation starts with an active Enable signal (EN), the bitline voltage differences "BL" and "BLB" are being detected by SA, and the output voltage is carried out. The suitable design of SA is the first objective for designing SRAM based circuits [9]. A CMOS inverter is attached additionally with SA output to amplify the magnitude. Due to high gain of inverter, the achieved Output (VOUT) and the cost of HSVLSA are high.

The Conventional Cross-coupled Voltage Latch SA (CCVLSA) was designed in 2019 [10] and its schematic is depicted in Fig. 1(b). Circuit consists of three sections: Driver, Access, and Load transistors. Various SAs for various memory cells are presented in literature [10]. VMSA is well-known because of its simplicity and performance. In voltage latch SA, EN needs to be selected carefully because output line also serves as input line. Footer Transistor M7 is used to bring down the leakage current during standby mode. When EN is high, then the SA conducts active mode.

The performance and current path in CCVLSA were improved by modifying it with a Double Switch Cross-coupled VMSA (DSCVMSA) [11] in 2013. For this purpose, an additional head switch and correspondent Enable (ENB) signal are introduced, as shown in Fig. 1(c). The head switch limits the weal latching transistors from being active during idle period, therefore the infirm current path is restricted, EN must be activated before ENB to transistor M1, and M3 remains to be turned "OFF". The Enable Bar signal turns "ON" the transistors M7 and M8. The output nodes charge equally to voltage of BL and BLB. When EN is active and ENB is inactive, the access transistors M7–M8 turn OFF. Then, node at the source of head switch M1 is picked up to Vdd and node at drain of bottom switch M6 takes out to the ground level. Thus, the output towards BL will be identical to power supply (Vdd) and output towards BLB will be equal to zero. This demonstrates that stored data in BL is 1 and stored data in BLB is 0.

A nine transistor Conventional Latched SA [12] is illustrated in Fig. 1(d). Here, transistors M2, M3, M5 and M6 form cross-coupled inverter. When EN is "1", PMOS transistors M8 and M9 are used for decoupling. M1 is precharge to supply voltage [8]. In CLSA, EN signal must be direct attentively due to difference in voltage between drain and source of PMOS M9 and



(a) HSVLSA. (b) CCVLSA. (c) DSCVMSA.

(d) CLSA. (e) TGVLSA. (f) DTGVSA.

Fig. 1: Circuit diversity.

M8, and leakage of current flow from output node to bit lines.

The design for Transmission Gate Voltage Latch SA (TGVLSA) was reported in 2019 [10] and its schematic is depicted in Fig. 1(e). The design consists of four PMOS (M6, M8, M1 and M2) and five NMOS (M3, M4, M5, M7 and M9) transistors. TGVLSA works in the same way as CCVLSA, only two additional transistors M5 and M7 are added by two PMOS transistors M6 and M8 which form a transmission gate. The transmission gate transistors are attached to BLB and BL. The delay and power of TGVLSA are improved as compared to CCVLSA. Therefore, transmission gate is the combination of PMOS and NMOS also known as analog or CMOS switch. TGVLSA is used to allow or reject a signal towards output [13] and [14]. The transmission gate is used to control gate signal and acts as a bi-directional switch.

The Dual switch Transmission Gate Voltage latch SA (DTG-VSA) [11] was designed to depreciate the power consumption and remove the leakage current is depicted in Fig. 1(f). The head switch transistor M10 and foot switch transistor M10 relates to enable EN respectively in the SA to avoid leakage current. Sleep switches are used to decrease leakage current [15]. The SA comprises of five PMOS and four NMOS transistors where transistors M5–M8 are used to make inverter circuit and transistors M3 and M4 are access transistors. When EN is "high" and ENB is "low", the sleep transistors M9 and M10 are turned "ON". Assuming BL is "high" and BLB is "low". When EN is equal to 0 and ENB is 1, M9 and M10 gets turned "OFF" and transmission gate transistors (M1, M3, M4, M2) will be turned "ON". By this, output node charges BL and BLB equally. Due to dual switch and decreased leakage current, the energy is falling off and the speed is expanded.

3. Schematic Diversity and Parameter Extraction of the Proposed Sense Amplifier

Differential Sense Amplifiers are used as the static current is not passed through them which will help in reducing power. Also, they are authentic. The main liability of SA is to detect the stored data of bitline voltage discharge. However, differential SA uses most silicon area. Therefore, high packing density is required. A Level Restoration Circuit (LRC) sets out a voltage swing from "0" to "1" logic level. LRCs can be used in VLSI Metal-Oxide-Semiconductor (MOS)

digital circuit designs. In Fig.2, primary level restoration circuit along with pass transistor is presented [16]. LR circuits are normally brought in because of their authenticity and also have low short circuit dissipation due to positive feedback transistor (M2).

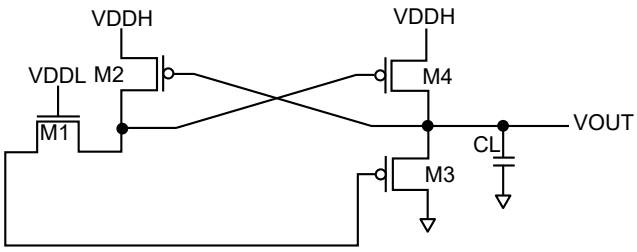


Fig. 2: Primary level restoration circuit along pass transistor.

3.1. Schematic Design of the Proposed Circuit

A Dual-Voltage Dual-Tail Level Restoration Voltage Latch Sense Amplifier (DVDTLR-VLSA) including two sleep transistors is reported in the paper. The schematic design of the proposed SA consists of ten transistors, as depicted in Fig. 3. The circuit design of the proposed SA is the same as double switch cross coupled VMSA [11], only two additional PMOS transistors M1 and M2 are attached. Also, there are two divisions of circuit connected to Vdd. The DVDTLR-VLSA amplifies the bitline voltage difference ΔVBLB and ΔVBL.

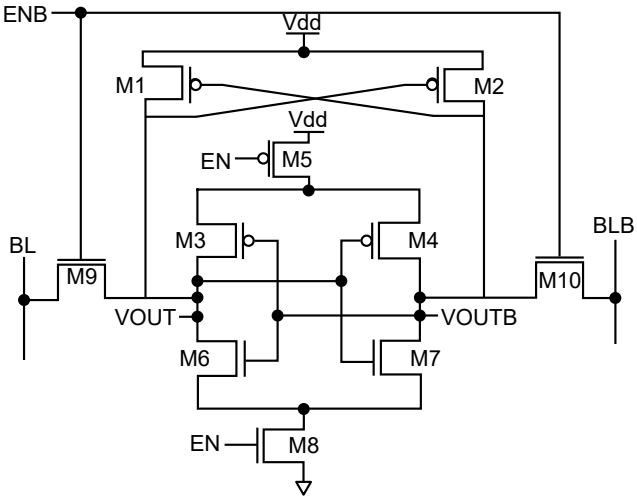


Fig. 3: Schematic diagram of the proposed SA.

Level restoration circuits are used in MOS logic designs as they supply low-to-high valid logic level of voltage swing in reaction to a low swing input signal Dual voltage with swing restoration logic is used [16]. When EN is "1" and ENB is "0", the enabling transistors M8 and M5 are turned "ON". Assuming (BLB = "0") and (BL = "1"), while ENB = 1, EN = 0, and sleeping transistor gets turned "OFF". This will charge output nodes.

3.2. Performance Analysis of the Proposed SA

In this section, the simulated results of the proposed SA are examined. All the simulations have been performed using LT SPICE tool. In Fig. 4(a), output functionality of the proposed SA is presented. When EN signal and BL are high, then output $V(p)$ follows the BL. Similarly, when ENB is active low and BL is high, the output $V(q)$ acts in accordance with BLB. By this waveform, working of SA can be confirmed. In Fig. 4(b) and Fig. 4(c), power and current waveforms are presented respectively.

Power is a critical parameter nowadays. The low power is important to avoid supplementary system required for cooling which can cause a portable problem [17] and [18]. Thus, cost of the systems raises due to additive system.

3.3. Dimensional Analysis of the Proposed SA

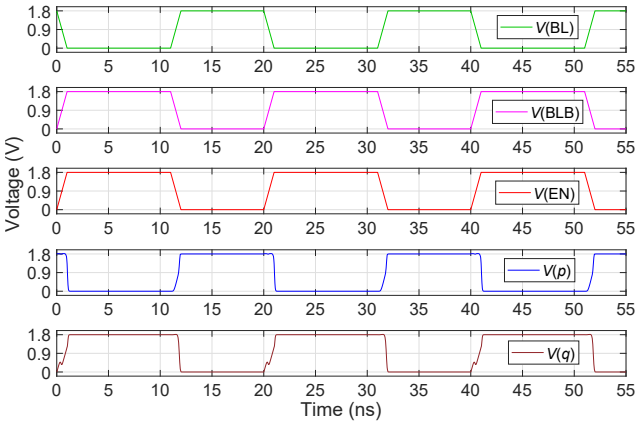
Speed performance of any SA can be recognized by knowing how fast the output nodes become charged. Therefore, correct designing is necessary to charge output nodes through precharge transistors in minimum time.

In Fig. 5, the aspect ratio of driver transistors and load transistors is arranged to attain perfect pull-up ratio and cell ratio [19]. Length of all driver and load transistors is 180 nm, while the widths (w) of transistors M3, M4, M6 and M7 are 900 nm, 900 nm, 300 nm and 300 nm respectively.

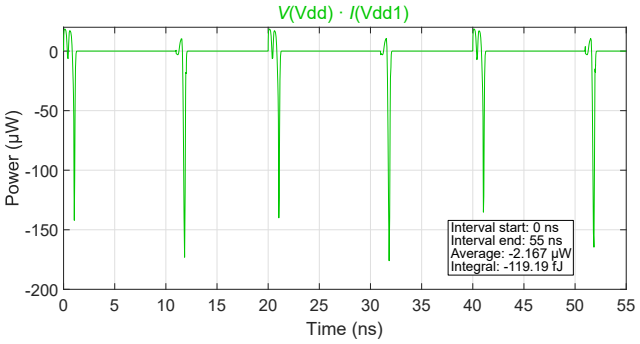
The design is analysed for low-power, low-energy, and satisfactory speed operation. The proposed SA is designed to store and amplify low voltage difference between the bitlines to yield desired logic level and also to reduce time for charging the output node. It can be surmised from Fig. 6(a) and Fig. 6(b) that when the aspect ratios of driver and load transistor are 900/180 nm and 300/180 nm, the best output results are achieved by this ratio as the output node charges fast. By this method, the speed of SA can be improved.

3.4. Sleep Transistor Effect on the Proposed SA

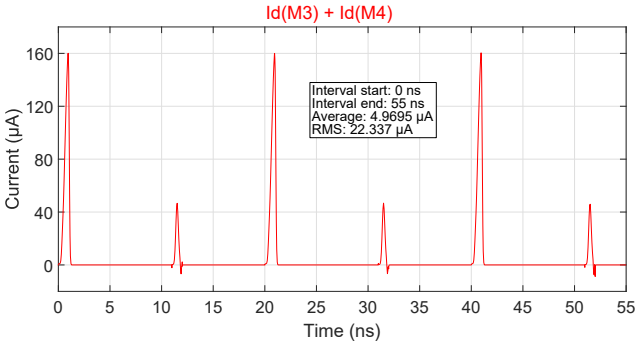
One among the low-power techniques used for VLSI circuits is power gating. It is used to save the power



(a) Functionality analysis.



(b) Power analysis.



(c) Current analysis.

Fig. 4: Output analysis of SA.

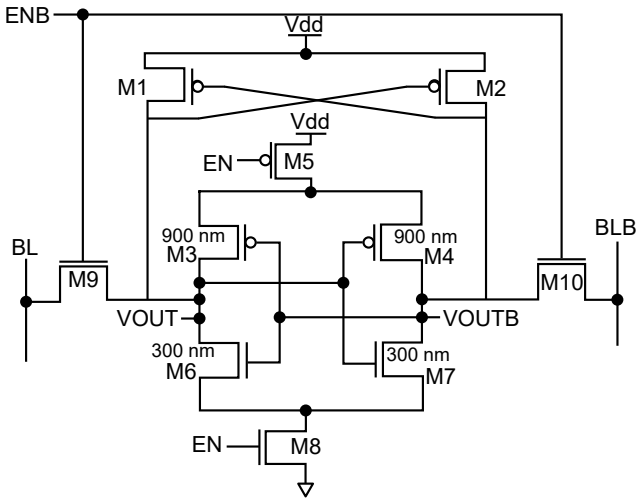
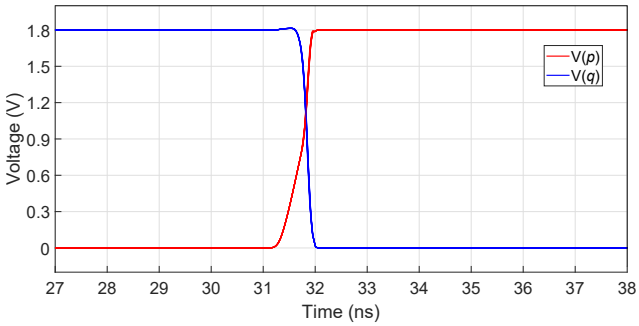
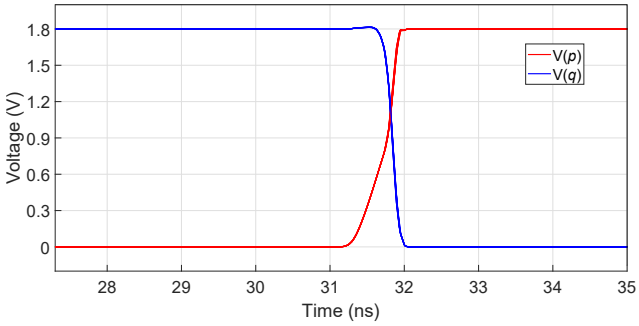


Fig. 5: Schematic of the proposed SA with labelled aspect ratios.



(a) PMOS $w = 900$ nm.



(b) NMOS $w = 300$ nm.

Fig. 6: Output waveform with various width.

by turning off the switch of the unused portion of circuit. PMOS and NMOS transistors are used as power switches. Another name for power switches is sleep transistors. NMOS and PMOS are used as power supply rails Vss and Vdd, respectively.

The impact of sleep transistor on the SRAM is reported by researchers in reference [20]. In Fig. 7, the bottom sleep transistor is removed from the proposed SA. After removing bottom sleep transistor, we get power and delay of 3.7 µW and 256.78 ps, respectively. Table 1 shows the result comparison of SA with or without sleep transistor.

Tab. 1: Effect of sleep transistor.

	Power (µW)	**Delay (ps)**
Proposed SA	2.16	262.03
Proposed SA without sleep transistor	3.70	256.78

It is shown in Tab. 1 that SA with sleep transistor is consuming lower power i.e., 2.16 µW than that of SA without sleep transistors i.e., 3.7 µW while there is not too much variation in delay results. It is due to reduction of leakage. The NMOS transistor is at-

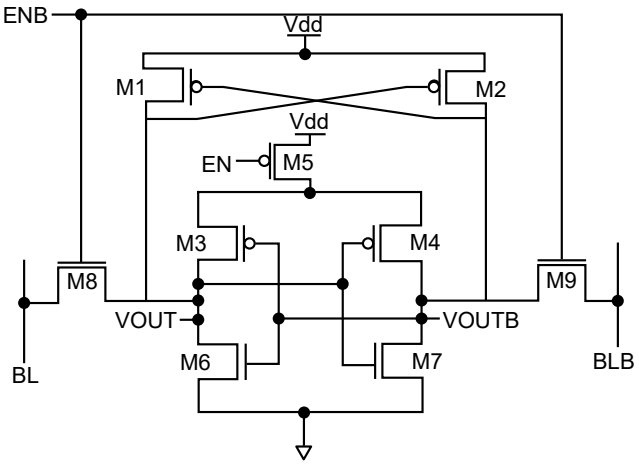


Fig. 7: Proposed SA without bottom sleep transistor.

tached to circuit and ground. Hence, it is called footer transistor.

4. Performance Comparison of Existing SA with the Proposed SA

Comparative performance analysis of various cross coupled voltage mode SAs with the proposed SA is presented in this section. The four main designs aimed in the VLSI industry are speed, delay, power, and area. Paper is summed up based on the design extent to analyze the stability within these parameters [20] and [21]. Low power is necessary for portable devices [22].

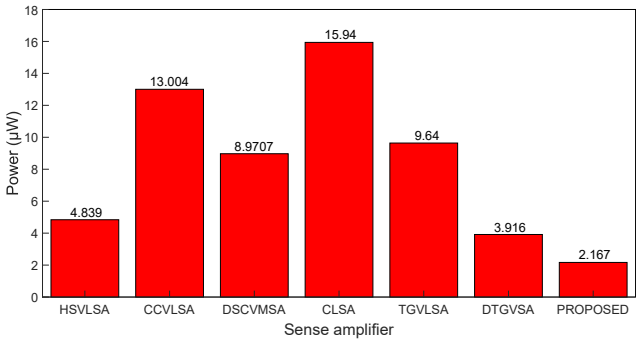
4.1. Energy and Power Performance

Power consumption is a principal parameter of entire VLSI circuits. With the rising request for low power circuits, research on lowering power consumption for SA is getting popular [23]. It may be surmised from Fig. 8(a) that among existing SAs, DVDTLR-VLSA has the low energy as well as the lowest power. Comparative energy analysis is shown in Fig. 8(b). The product of delay and power also known as switching energy is used to calculate the energy. Here, Power $P = V \cdot I$.

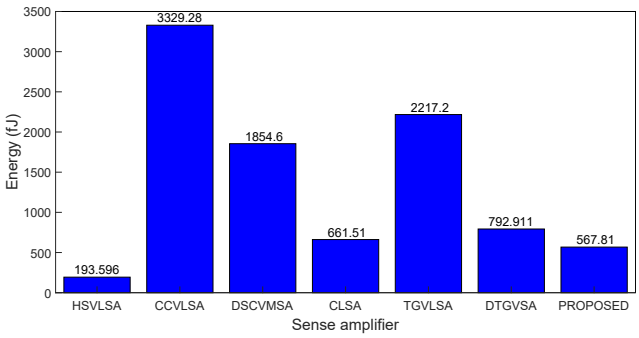
4.2. Delay Analysis

As the interest of high-speed circuit increases, delay factor becomes more important in VLSI circuits [24].

It can be surmised from Fig. 9 that SA i.e., HSVLSA has the lowest delay and therefore, it is the best choice. Also, there is a tradeoff between other topologies pre-



(a) Analytical output waveform of comparative power.



(b) Analytical output waveform of comparative energy.

Fig. 8: Analytical output waveform of (a) comparative power and (b) comparative energy.

sented in Tab. 2. Propagation delay of logic circuit is defined as time needed for output voltage to hold out the midway between the 0 and 1 logic levels. i.e., half (50 %) of Vdd:

$$\tau_p = \frac{\tau_{phl} + \tau_{plh}}{2}, \tag{1}$$

where τ_p, τ_{phl} and τ_{plh} are known as total propagation delay, high-to-low transition delay and low-to-high transition delay, respectively. Results are obtained by assuming aspect ratios of PMOS's of 4 : 1 and NMOS's of 2 : 1.

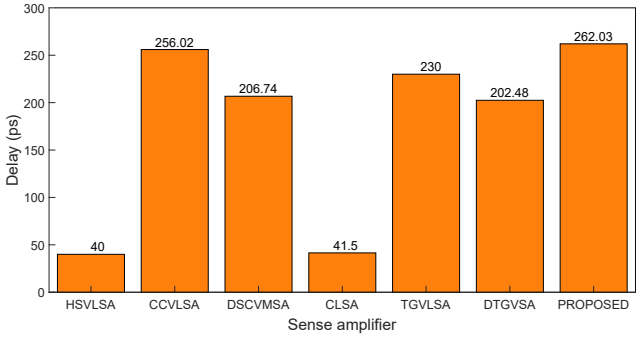


Fig. 9: Delay comparison.

Tab. 2: Comparison of SA configuration designed at 180 nm technology node and evaluated for 1.8 V supply [10].

Sense amplifiers with power supply 1.8 V	Power (µW)	Energy (fJ)	Delay (ps)	Average current (µA)	Number of transistors
HSVLSA	4.839	**193.596**	**40**	0.412	7
CCVLSA	13.004	3329.28	256.02	0.796	7
DSCVMSA	8.9707	1854.60	206.74	0.514	8
CLSA	15.94	661.51	41.50	2.230	9
TGVLSA	9.64	2217.2	230.00	2.504	9
DTGVSA	3.916	792.911	202.48	2.229	10
Proposed SA	**2.167**	567.81	262.03	**4.878**	10

4.3. Comparative Current Analysis

In Fig. 10, comparative results of current of all SA topologies are presented. Current (I) equals to power divided by voltage ($I = P \cdot V^{-1}$). Here we have calculated the average current passing towards the output nodes.

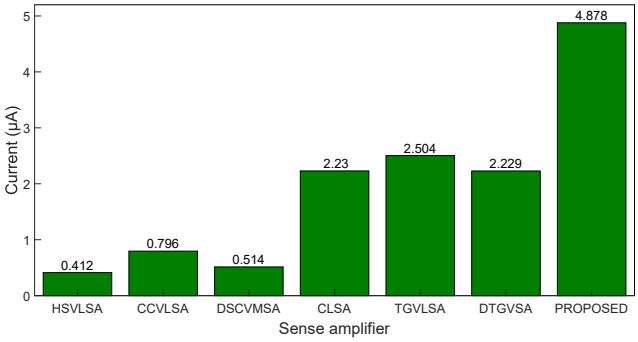


Fig. 10: Current analysis.

It can be surmised from the Fig. 10 that HSVLSA has the lowest current and DVDTLR-VLSA possesses the maximum current i.e., 4.878 µA. It is because of transistor count and dual voltage present. Current of the proposed design is high, but consuming lowest power. The comparative data of parameters; power, switching energy delay and current of each SA evaluated for 1.8 V is given in Tab. 2 [25].

5. Conclusion

This paper has decorated performance upgradation of Sense Amplifier (SA) and a dual-voltage dual-tail level restoration voltage latch SA (DVDTLR-VLSA) has been reported. The SA is designed at a 180 nm technology node and 1.8 V power supply. Comparative Transient analysis for each SA with the proposed SA is done at 55 ns pulse width. The DVDTLR-VLSA is modified mainly for high-speed and low-power consequences. It can be inferred from the simulated results that the power of the modified SA is the lowest, specifically 2.2 µW. That is notably improved compared to existing SAs analysed in the paper. Two additional analyses have been done for DVDTLR-VLSA which show the proposed design to be the best case. As in the case of SA without sleep transistors, the power is increased. The DVDTLR-VLSA is designed for low-power circuits.

Author Contributions

D.D. and P.M. conceived the idea and planned the complete work. D.D. and B.R. has carried out all the simulations and plotted/analysed all the results. D.D. and P.M. performed the calculations and analysed the data for interpretation of the results and graph modification. P.M. and B.K. designed the model and the computational framework. D.D. and B.R. drafted first write-up of the manuscript. B.R., P.M. and B.K. prepared the revised manuscripts and prepared the detailed response to the reviewer comments. All the authors provided critical feedback and helped shape the research, analysis and manuscript.

References

[1] RAWAT, B. and P. MITTAL. Single bit line accessed high-performance ultra-low voltage operating 7T static random access memory cell with improved read stability. *International Journal of Circuit Theory and Applications*. 2021, vol. 49, iss. 5, pp. 1435–1449. ISSN 1097-007X. DOI: 10.1002/cta.2960.

[2] ZHU, J., N. BAI. and J. WU. A Review of Sense Amplifiers for Static Random Access Memory. *IETE Technical Review*. 2013, vol. 30, iss. 1, pp. 72–81. ISSN 0974-5971. DOI: 10.4103/0256-4602.107343.

[3] NEGI, S., P. MITTAL and B. KUMAR. Modeling and Analysis of High-Performance Triple Hole Block Layer Organic LED Based Light Sensor for Detection of Ovarian Cancer. *IEEE Transactions on Circuits and Systems I: Regular Papers*. 2021, vol. 68, iss. 8, pp. 3254–3264. ISSN 1558-0806. DOI: 10.1109/TCSI.2021.3078510.

[4] GARG, M. R. and A. TONK. A Study of Different Types of Voltage & Current Sense Amplifiers used in SRAM. *International Journal of Advanced Research in Computer and Communication Engineering*. 2015, vol. 4, iss. 5, pp. 30–35. ISSN 2278-1021. DOI: 10.17148/IJARCCE.2015.4507.

[5] KOBAYASHI, T., K. NOGAMI, T. SHIROTORI and Y. FUJIMOTO. A current-controlled latch sense amplifier and a static power-saving input buffer for low-power architecture. *IEEE Journal of Solid-State Circuits*. 1993, vol. 28, iss. 4, pp. 523–527. ISSN 1558-173X. DOI: 10.1109/4.210039.

[6] JAIN, R., K. GUPTA. and N. PANDEY. Hybrid Dynamic CML with Modified Current Source (H-MDyCML): A Low-Power Dynamic MCML Style. *Advances in Electrical and Electronic Engineering*. 2021, vol. 19, iss. 1, pp. 57–65. ISSN 1804-3119. DOI: 10.15598/aeee.v19i1.3944.

[7] RAWAT, B. and P. MITTAL. A 32 nm single-ended single-port 7T static random access memory for low power utilization. *Semiconductor Science and Technology*. 2021, vol. 36, iss. 9, pp. 6–22. ISSN 1361-6641. DOI: 10.1088/1361-6641/ac07c8.

[8] PATHRIKAR, A. K. and R. S. DESHPANDE. Design of faster & power efficient sense amplifier using VLSI technology. In: *2016 IEEE International Conference on Advances in Electronics, Communication and Computer Technology (ICAECCT)*. Pune: IEEE, 2016, pp. 358–361. ISBN 978-1-5090-3662-2. DOI: 10.1109/ICAECCT.2016.7942613.

[9] CHRISANTHOPOULOS, A., Y. MOISIADIS, Y. TSIATOUHAS and A. ARAPOYANNI. Comparative study of different current mode sense amplifiers in submicron CMOS technology. *IEE Proceedings - Circuits, Devices and Systems*. 2002, vol. 149, iss. 3, pp. 154–158. ISSN 1359-7000. DOI: 10.1049/ip-cds:20020425.

[10] KISHORE, C. and A. KUMAR. Comparative Analysis of Various Sense Amplifiers in 45nm CMOS Technology. *International Research Journal of Engineering and Technology*. 2019, vol. 6, iss. 5, pp. 102–106. ISSN 2395-0056.

[11] NA, T., S.-H. WOO, J. KIM, H. JEONG and S.-O. JUNG. Comparative Study of Various Latch-Type Sense Amplifiers. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*. 2014, vol. 22, iss. 2, pp. 425–429. ISSN 1557-9999. DOI: 10.1109/TVLSI.2013.2239320.

[12] GUNDU, A. K., M. S. HASHMI and A. GROVER. A New Sense Amplifier Topology with Improved Performance for High Speed SRAM Applications. In: *2016 29th International Conference on VLSI Design and 2016 15th International Conference on Embedded Systems (VLSID)*. Kolkata: IEEE, 2016, pp. 185–190. ISBN 978-1-4673-8700-2. DOI: 10.1109/VLSID.2016.38.

[13] KUMAR, N., P. MITTAL and M. MITTAL. Performance Analysis of FinFET based 2:1 Multiplexers for Low Power Application. In: *2020 IEEE Students Conference on Engineering & Systems (SCES)*. Prayagraj: IEEE, 2020, pp. 1–5. ISBN 978-1-7281-9339-7. DOI: 10.1109/SCES50439.2020.9236770.

[14] ETEZADI, S. and S. A. HOSSEINI. Novel Ternary Logic Gates Design in Nanoelectronics. *Advances in Electrical and Electronic Engineering*. 2019, vol. 17, iss. 3, pp. 294–305. ISSN 1804-3119. DOI: 10.15598/aeee.v17i3.3156.

[15] KUMAR, N., P. MITTAL, B. RAWAT and M. MITTAL. Dynamic Power Consumption and Delay Analysis for Ultra-Low Power 2 to 1 Multiplexer Designs. *Advances in Electrical and Electronic Engineering*. 2021, vol. 19, iss. 2, pp. 145–154. ISSN 1804-3119. DOI: 10.15598/aeee.v19i2.3821.

[16] MOISIADIS, Y., I. BOURAS and A. ARAPOYANNI. High performance level restoration circuits for low-power reduced-swing interconnect schemes. In: *7th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*. Jounieh: IEEE, 2000, pp. 619–622. ISBN 0-7803-6542-9. DOI: 10.1109/ICECS.2000.911615.

[17] PAL, P. K. and R. K. NAGARIA. A Low Power Low Supply MOS-Only Subthreshold Voltage Reference for Wide Temperature Range. *Advances in Electrical and Electronic Engineering*. 2019, vol. 17, iss. 2, pp. 167–178. ISSN 1804-3119. DOI: 10.15598/aeee.v17i2.3197.

[18] RATHOD, A. P. S., P. LAKHERA, A. K. BALIGA, P. MITTAL and B. KUMAR. Performance comparison of pass transistor and CMOS logic configuration based de-multiplexers. In: *International Conference on Computing, Communication & Automation*. Greater Noida: IEEE, 2015, pp. 1433–1437. ISBN 978-1-4799-8890-7. DOI: 10.1109/CCAA.2015.7148606.

[19] KANIKA and P. K. DAHIYA. Analyzing the Impact of Sleep Transistor on SRAM. *IOSR Journal of VLSI and Signal Processing*. 2017, vol. 7, iss. 3, pp. 64–67. ISSN 2319-4200. DOI: 10.9790/4200-0703016467.

[20] GOYAL, A., A. TOMAR, and A. GOYAL. Effect of W/L Ratio on SRAM Cell SNM for High-

Speed Application. *International Research Journal of Engineering and Technology*. 2017, vol. 4, iss. 4, pp. 2326–2332. ISSN 2395-0056.

[21] YANG, B.-D. and L.-S. KIM. A low-power SRAM using hierarchical bit line and local sense amplifiers. *IEEE Journal of Solid-State Circuits*. 2005, vol. 40, iss. 6, pp. 1366–1376. ISSN 1558-173X. DOI: 10.1109/JSSC.2005.848032.

[22] SHARMA, V., S. COSEMANS, M. ASHOUEI, J. HUISKEN, F. CATTHOOR and W. DEHAENE. A 4.4 pJ/Access 80 MHz, 128 kbit Variability Resilient SRAM With Multi-Sized Sense Amplifier Redundancy. *IEEE Journal of Solid-State Circuits*. 2011, vol. 46, iss. 10, pp. 2416–2430. ISSN 1558-173X. DOI: 10.1109/JSSC.2011.2159056.

[23] SEEVINCK, E., P. J. VAN BEERS and H. ONTROP. Current-mode techniques for high-speed VLSI circuits with application to current sense amplifier for CMOS SRAM's. *IEEE Journal of Solid-State Circuits*. 1991, vol. 26, iss. 4, pp. 525–536. ISSN 1558-173X. DOI: 10.1109/4.75050.

[24] BLALOCK, T. N. and R. C. JAEGER. A high-speed clamped bit-line current-mode sense amplifier. *IEEE Journal of Solid-State Circuits*. 1991, vol. 26, iss. 4, pp. 542–548. ISSN 1558-173X. DOI: 10.1109/4.75052.

[25] KHELLAH, M. M. and M. I. ELMASRY. A low-power high-performance current-mode multiport SRAM. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*. 2001, vol. 9, iss. 5, pp. 590–598. ISSN 1557-9999. DOI: 10.1109/92.953493.

About Authors

Divya DUTT was born in Delhi, India in 1997. She received her B.Tech. degree in Electronics and Communication Engineering from Raj Kumar Goel Institute of Technology, Ghaziabad, India in 2019. She received her M.Tech. degree from Delhi Technological University, New Delhi, India. Her research interests include Low Power VLSI Circuits, Amplifier Circuits.

Poornima MITTAL (IEEE member) has published 10+ international journal/conference papers/book chapters. Her research interest includes design/modelling of flexible electronic devices, material synthesis and characterization, electronic and optical materials, thin film fabrication, Organic Light-Emitting Diode (OLED), solar cell, memory design and low power VLSI circuits. She has published one patent on novel Organic Thin-Film Transistor structure and textbook titled "Organic Thin-Film-Transistor Applications: Materials to Circuits", CRC Press, United Kingdom, 2016. She is the reviewer of many IEEE transactions and other international journals of IEEE, Elsevier, Springer, IOP, AVS, Wiley and T&F. She has received the research awards in 2012 and 2015 for dedicated research at Graphic Era University, Dehradun, India. Also, she has received Commendable- Research-Award in 2019, 2020, 2021 and 2022 at Delhi Technological University (DTU), Delhi, India. She is the life member of many professional societies. She has more than 15 years of academic and research experience and presently, she is working as professor in the department of Electronics and Communication Engineering at DTU, Delhi, India.

Bhawna RAWAT was born in Delhi, India in 1993. She received her B.Tech. degree in Electronics and Communication Engineering from Shiv Nadar University, Greater Noida, India in 2015. She received M.Tech. degree in VLSI Design from Indira Gandhi Delhi Technical University for Women, Delhi, India. She is presently pursuing Ph.D. at Delhi Technological University, New Delhi, India. Her research interests include low power circuit design, memory circuits in emerging technologies.

Brijesh KUMAR is presently working as a professor in the Department of Information Technology, Indira Gandhi Delhi Technical University for Women, Delhi, India since April 2022. Prof. Kumar earned a B.Tech. degree at Bundelkhand Institute Engineering and Technology (B.I.E.T.), Jhansi (Uttar Pradesh) India; and M.Tech. at Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India; and a Ph.D. at Indian Institute of Technology (IIT), Roorkee, India. He has more than 150 international, national journals and conference publications, patents, book, book chapter publications as well. His research interest includes Internet of Things (IoT) Based Smart Devices and Testing, Artificial Intelligence and Machine Learning Applications, Data Communication and Computer Network, Memory Designs, Flexible Electronics, Electronic Devices and Circuits, VLSI Design and Technology, Digital System Design, Semiconductor Novel Device Structures.

Food Engineering Reviews
https://doi.org/10.1007/s12393-022-09323-1

Designs, Performance and Economic Feasibility of Domestic Solar Dryers

Shimpy[1] · **Mahesh Kumar**[1] · **Anil Kumar**[2,3]

Received: 10 March 2022 / Accepted: 29 September 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Present article is an overview of available solar drying technologies developed for small rural agricultural farms emphasizing domestic applications. A huge amount (about 61%) of perishable items gets wasted annually at the household level due to lack of awareness, negligence, improper handling, and inadequate storage facilities. Domestic solar dryers are reviewed and presented under the categories of natural and forced convection modes. The maximum attainable temperature inside the drying chamber under natural and forced convection mode is observed 98.6 and 78.1 °C, respectively. Thermal efficiency of solar dryers varies from 5.16 to 64.36% for the drying of various commodities. Natural convection solar dryers are appropriate for rural and undeveloped areas due to simple design and lower capital and electrical requirements. In comparison, forced convection solar dryers are more attractive due to better performance, higher drying rate, and lower drying time for high moisture content products. The designs of indirect and mixed-mode solar dryers seem very rare in the area of domestic solar drying. Solar dryers have potential to reduce the conventional drying cost by 50% and improve the return by 30%.

Keywords Solar drying · Domestic solar dryers · Natural convection · Forced convection · Economics

Introduction

Electricity is still a dream in many places globally and is also very expensive for most developing nations. Developed countries with good electric supply are also concerned about the limited fossil fuel reserves and the pollution caused by them in electricity generation. Hence, using solar energy for various applications is an effort of global importance. It eliminates the requirements of conventional fuels without causing any kind of environmental pollution [1].

Sun is the ultimate source of energy on the earth. There are different technologies to harness the available solar energy. Open sun drying (OSD) is one of the techniques used traditionally for food preservation. It has some severe limitations, such as uncontrolled operating conditions, insect and animal interference, etc., which reduces the product quality and affects the quantity of the dried product. Despite all the drawbacks, OSD is still practiced by many farmers and households around the globe for drying different commodities such as fruits, vegetables, spices, grains, medicinal herbs, aromatic plants, wood, etc. [2]. The moisture content present in a food product is highly responsible for its spoilage or shorter shelf-life. Higher the amount of moisture content, higher the chances of deterioration of the food product. In solar drying, products are placed in a warm environment to remove moisture until it reaches a safe moisture content of about 8–10% (wet basis), which is required for its "safe storage time period" during which the quality of the products remains unaffected from any microbial growth [1, 2].

A total food loss and waste of about 20% of oilseeds, meat and dairy products, 35% of fish, 30% of cereals, and 40–50% of horticultural crops annually have been reported [3]. Household sector contributes 61% to the total global food waste, followed by the service and retail sectors with 26% and 13%, respectively [4]. Food loss in India is estimated to be 10.6 billion USD [5]. Stanley and Colo developed a solar dryer in 1976 to reduce food losses and improve the quality

✉ Mahesh Kumar
mkshandilya1@gmail.com

Shimpy
shimpymehra@gmail.com

Anil Kumar
anilkumar76@dtu.ac.in

[1] Department of Mechanical Engineering, Guru Jambheshwar University of Science & Technology, Hisar, India

[2] Department of Mechanical Engineering, Delhi Technological University, Delhi, India

[3] Centre for Energy and Environment, Delhi Technological University, Delhi, India

Published online: 20 October 2022

of the product [6]. Since then, considerable research has been done on solar energy utilization in various solar dryers. Solar dryers available in the literature can be classified based on modes of heat transfer to the drying commodity and air circulation through the system (Fig. 1).

Solar drying is a two-step process; the first step is 'heat transfer' from the Sun to the drying commodity in the form of solar radiation and the second step is 'mass transfer' from the drying commodity to the surrounding in the form of evaporated moisture. In this way, a solar dryer can be classified firstly based on the mode of heat transfer as a direct, indirect and mixed-mode type solar dryer [8] and secondly based on the mode of mass transfer as a natural and forced convection solar dryer [9].

The drying commodity is placed over the drying tray inside a drying chamber with transparent walls/roofs in a direct solar dryer. After passing through the transparent walls/roof, solar radiations directly strike the product and get absorbed, causing temperature rise and moisture evaporation. Some drawbacks are also associated with this process due to direct contact between solar radiation and product to be dried. Direct exposure to sunlight may lead to discoloration [10] and loss of some vitamins in the food items [11], which affects their quality. A direct solar dryer can be a cabinet or tunnel type [12]. In an indirect type solar dryer, the drying commodity is placed in a well-insulated drying chamber to avoid direct exposure to solar radiation. Drying air is heated separately using solar air heaters and then supplied to the drying chamber. The involvement of solar air heaters makes the system more complicated than direct solar dryers. However, the product quality obtained from indirect solar dryers is much better. The mixed-mode solar dryer combines both solar air heating actions, i.e., direct and indirect solar air heating. Drying chamber comprises transparent walls/roof, and a solar collector arrangement is used to obtain higher temperatures with better air circulation [13].

Solar-heated air circulates through the drying chamber in a natural convection solar dryer due to buoyancy forces. These dryers are called "passive type" or "free convection" solar dryers. This type of dryer needs no electric supply or other power to work except solar radiation, making them more attractive to rural or remote areas where electricity is unavailable or costs too high. The lower drying rate is one of the major limitations of this type of dryer [1]. A forced convection solar dryer, on the other hand, uses an additional device such as a fan, pump, blower, etc., to circulate the solar-heated air through the drying chamber. These dryers are also known as "active solar dryers". Drying rates are higher for these dryers than the natural convection solar dryers. Still, there is a need for electrical energy to drive the pump or blower, which makes them more complex, dependent on conventional energy, and increases the system's total cost [14].

Developing simple, compact-sized, easily operated, low-cost solar drying systems is very attractive to households and small agricultural farm users [15]. These systems generally depend upon solar energy and are independent of electric supply. In this way, no capital investment is required for its operation, and there is no harmful emission in using them. The abundance of solar energy around the globe makes these systems equally important to the entire world. Solar dryers are preferred over other dryers for domestic applications [16]. Most solar drying systems' efficiency is reported in the range 30–50% [17]. Many researchers have reviewed solar dryers with various considerations [7, 18, 19].

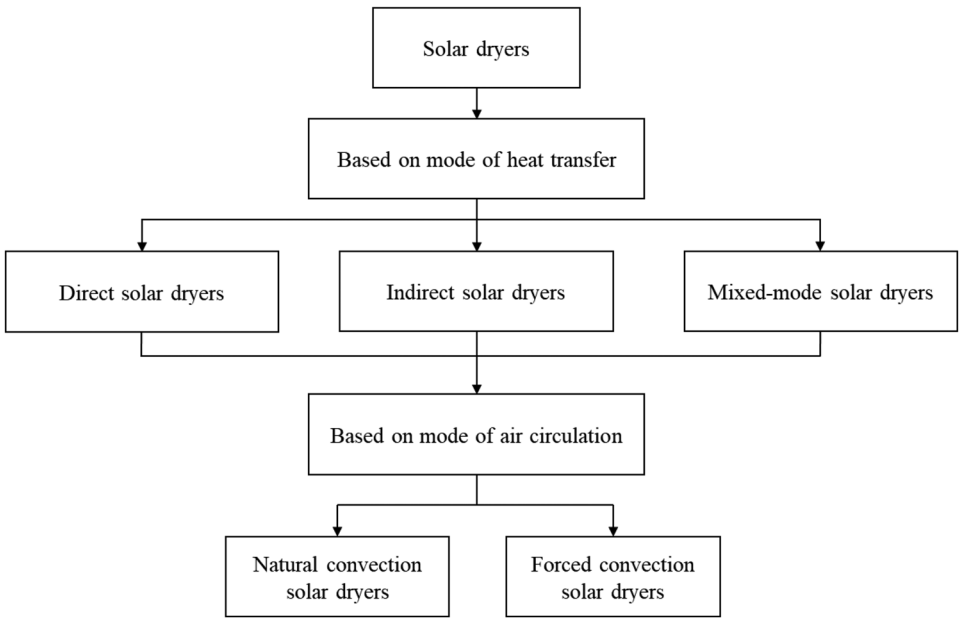


Fig. 1 Classification of solar dryers [7]

This communication presents a brief overview of the progress in low-cost, compact-sized solar dryers developed for domestic and small agricultural farm applications. A systematic review method was adopted for the present study. Firstly, the keywords such as solar dryer, domestic solar dryer, household solar drying, small-scale solar dryers, low-cost solar dryers and solar food drying, etc. were used to search papers from the literature available on different search engines such as Google Scholar, Springer Link, Science Direct, Wiley Online Library, etc. Around one hundred articles were collected from 1986 to 2021. Out of which 76 papers are discussed based on dryer design, significance of results, and reputation of journal/publisher. The comprehensive studies on domestic solar dryers have been categorized as domestic type natural convection solar dryers (NCSD) and forced convection solar dryers (FCSD). Construction and working details of various domestic solar dryers tested for different commodities have been presented. Economic perspective has also been given to understand the feasibility of various designs.

Natural Convection Solar Dryers (NCSD)

Natural convection solar dryers (NCSD) are simple in design, need lower capital investment, and are independent of electricity requirements. The given factors make NCSD more suitable for domestic, small agricultural farms, rural and undeveloped areas. NCSD can be used for different drying commodities but is generally tested for various household items such as fruits and vegetables with higher spoilage chances.

Investigations on Domestic NCSD for Fruits Drying

Fournier and Guinebault [20] tested a shell-type solar dryer (Fig. 2) to dry mango slices. Thickness of the fruit slice was suggested to be a critical parameter that significantly affected drying time. Hallak et al. [21] developed a staircase-shaped direct NCSD (Fig. 3) for drying apricots, figs, grapes, prune, okra and tomatoes, having efficiency in 26 to 65% under variable weather and ambient conditions.

Wakjira et al. [22] developed and tested an enclosed-cabinet solar dryer for optimum slice thickness of banana samples. Dryer took 4 days to attain safe moisture content for slices of 3–4 mm thickness with the best quality. Rawat et al. [23] fabricated an NCSD having a quadrilateral-shaped cabinet covered with a 2 mm thick transparent glass, 120×78 cm^2 cabinet floor area. The amount of fuel saved for drying 5 kg of amla (Phyllanthus emblica L.) from a moisture content of 74.4 to 2% wet basis (w.b.) in 3 days by using NCSD was calculated to be 550.46 kg, 176.71 L, 366.97 kg, and 141.14 kg in terms of fuel-wood, light diesel oil, coal and natural gas, respectively. Energy payback time ranges from 3.1–5.0 years for various conventional fuels. Terres et al. [24] developed a direct NCSD (Fig. 4) having two separate transparent glass covers for drying lemon slices. The maximum temperature in the middle of the drying chamber was 47.2 °C. ANSYS software was used to simulate the drying chamber for temperature, density and air currents.

Daud and Simate [25] used a solar conduction dryer (SCD) having polyethylene plastic glazing and four drying trays (surface area: 0.1131 m^2 each) for the drying of sliced pineapples. The maximum inside temperature (46 °C) was 9 °C higher

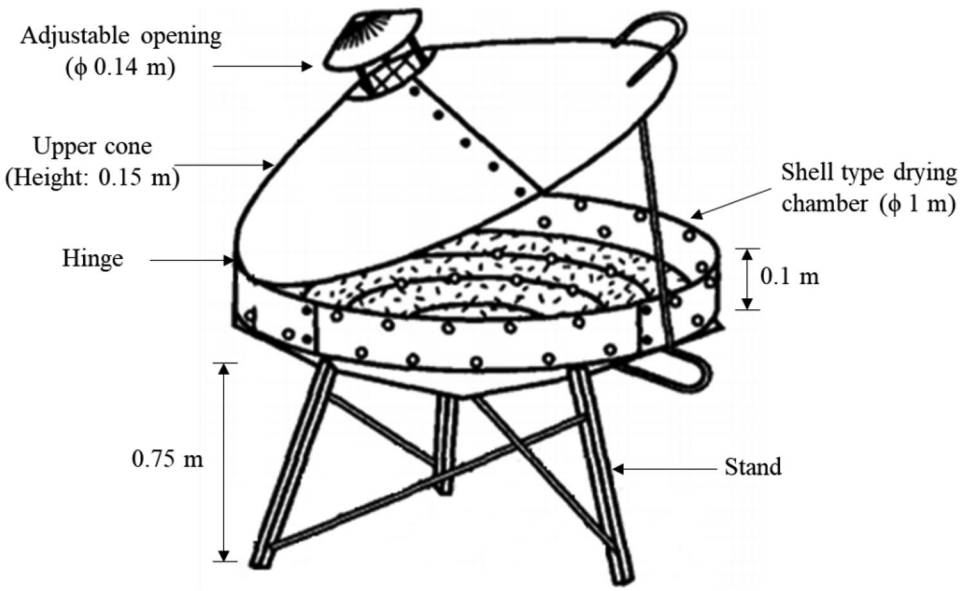


Fig. 2 Shell-type solar dryer [20]

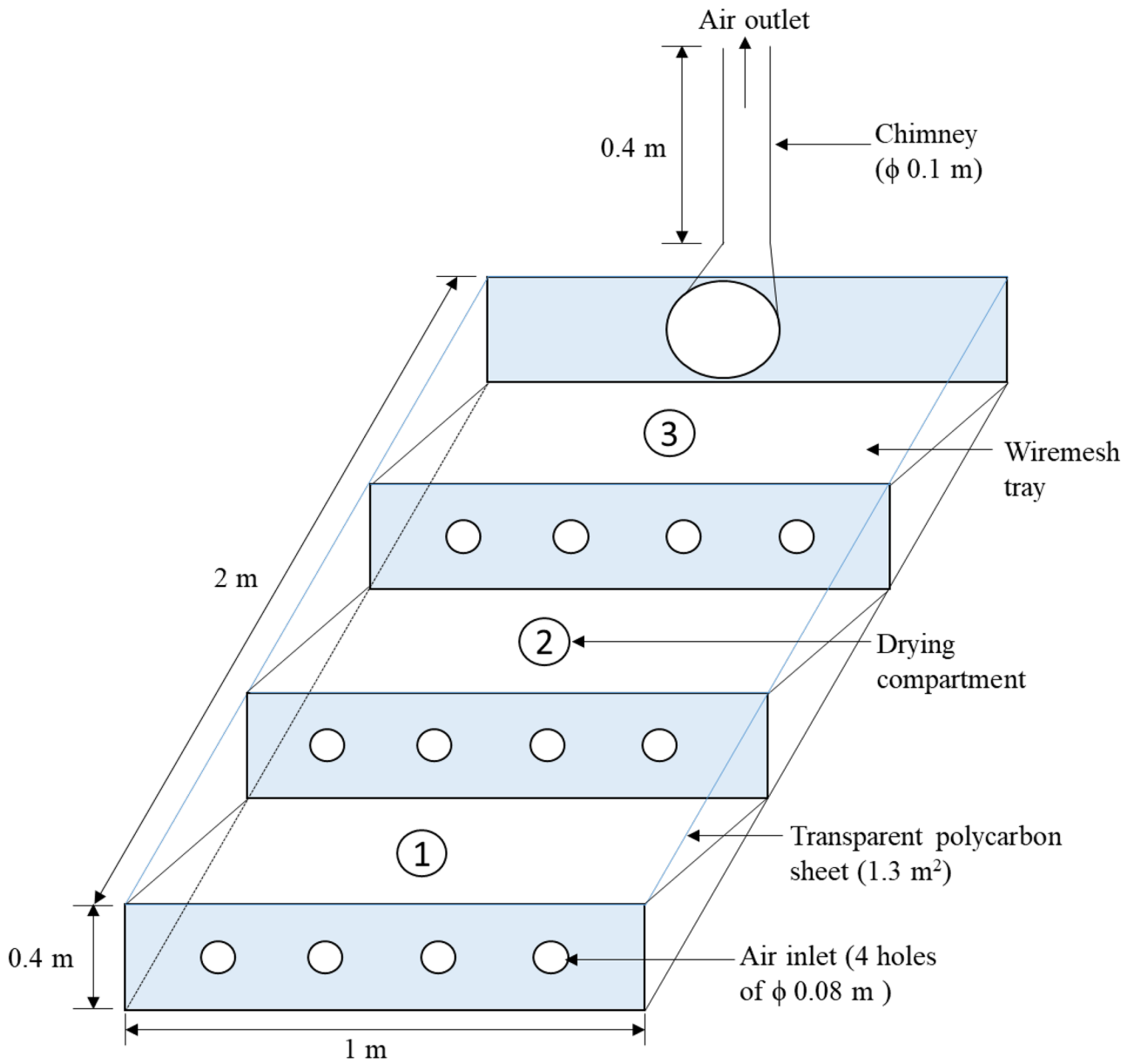


Fig. 3 Schematic of a staircase-shaped direct NCSD [21]

than the ambient temperature. Moisture content was reduced from 85.42 to 12.23% (w.b.) by using SCD in 8 h, whereas it was just up to 51.51% (w.b.) under OSD. Poonia et al. [26] developed and tested a simple, low-cost NCSD for drying Ziziphus (ber). Maximum temperatures of 63 °C and 49 °C were observed inside the drying chamber under no-load and loaded conditions, respectively. The moisture content of the ber sample was reduced from 85 to 15% (w.b.) with an overall efficiency of 16% and a total cost of $10.31. Islam et al. [27] constructed a direct NCSD (Fig. 5) with three separate drying chambers having different arrangements for ventilation (a thin tube chimney type, attic type and simple type) and tested for the drying of apple, banana and guava. Simple type drying chamber was the most efficient with a maximum moisture removal rate of 58.9% compared to 44.5% for the chimney type and 33.3% for the attic type ventilated drying chamber. Summary of various investigations carried out on domestic NCSD for fruit drying has been presented in Table 1.

Investigations on Domestic NCSD for Vegetable Drying

Sharma et al. [28] developed and tested a cabinet type direct NCSD (Fig. 6) to dry peas, grapes and potatoes. The maximum tray temperature went up to around 80 °C, about 50 °C

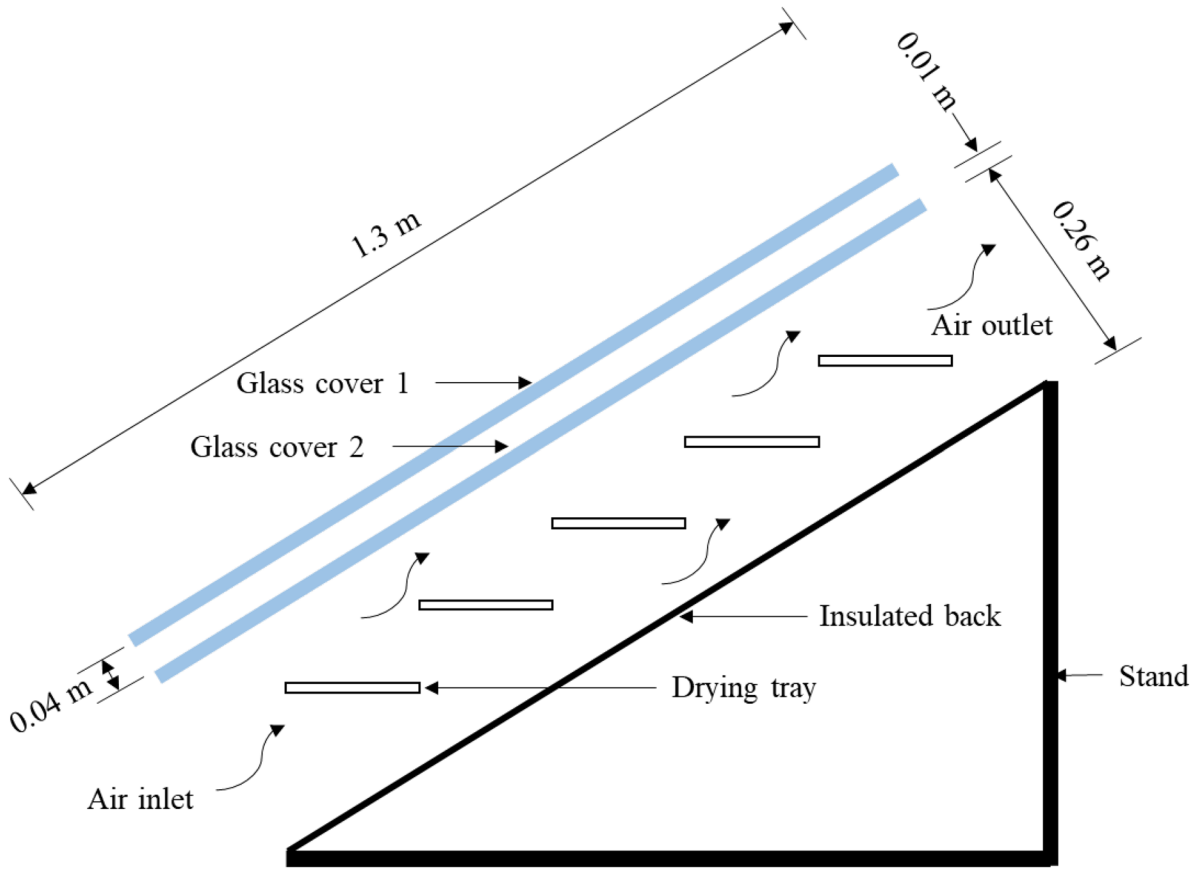


Fig. 4 Solar dryer for lemon slice drying [24]

higher than the ambient temperature. Thanvi and Pande [29] designed and tested a low-cost NCSD (Fig. 7) to dry chilies. Average inside temperature was 63.1 °C which is 36.4 °C higher than the ambient. Solar energy utilization efficiency was increased by 6.7% as compared to OSD. Sharma et al. [30] compared the performance of the cabinet type direct NCSD and OSD for the drying of turnip, cauliflower and green peas. The maximum temperature inside the drying chamber was found to be 85 °C and 50 °C under no-load and loaded conditions, respectively. Ezekoye and Enebe [31] designed an NCSD, as shown in Fig. 8 and tested it for drying agricultural products (pepper and groundnuts). The temperature at the drying trays varied from 33 to 67 °C for a variation of 25 to 31 °C in the ambient temperature.

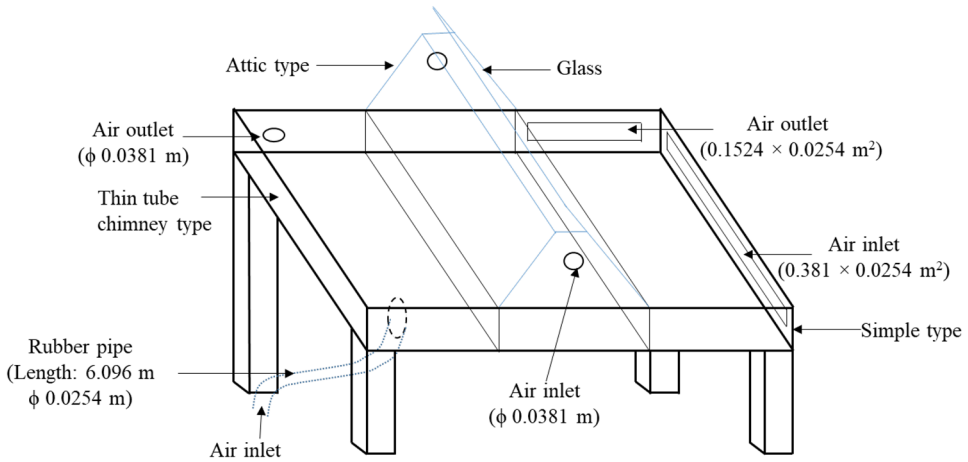


Fig. 5 Direct solar dryer with three ventilation arrangements [27]

Table 1 Summary of domestic NCSD for fruits drying

Sr. No.	Researcher/s	Year	Country	Drying commodity	Description	Remarks
1	Fournier and Guinebault [20]	1995	France	Mango slices	A theoretical model was developed for a shell-type solar dryer	Slice thickness and wind velocity were the most affecting parameters
2	Hallak et al. [21]	1996	Israel	Apricots, figs, grapes, prune, okra and tomatoes	A novel staircase-type solar dryer was developed and tested	A maximum airflow rate of about 1.5 m^3/min was reported inside the dryer
3	Wakjira et al. [22]	2011	Ethiopia	Banana slices	Enclosed-cabinet solar dryer was tested for optimum slice thickness and shelf-life estimation	Slice thickness of 3–4 mm was recommended for better quality products with longer shelf-life
4	Rawat et al. [23]	2014	India	Amla (Phyllanthus emblica L.)	Energetics and CO_2 mitigation potential of a laboratory-scale NCSD were reported	Annual CO_2 mitigation was 8,812.40 to 34,368.37 kg in terms of fuel-wood, light diesel oil, coal and natural gas
5	Terres et al. [24]	2015	Mexico City	Lemon slices	Direct NCSD having two separate glass covers and multiple trays was tested	Vorticity has been observed in some parts of the drying chamber
6	Daud and Simate [25]	2017	Zambia	Sliced Pineapples	Drying kinetics, the effect of slice thickness and the used drying tray area were investigated for SCD	Middilli model was the most suitable for describing the drying behavior. The highest drying rates were observed with 3 mm slice thickness. Maximum thermal efficiency (44%) was observed with 100% drying tray area utilization
7	Poonia et al. [26]	2018	India	Ber (Ziziphus)	Simple low-cost NCSD was fabricated and tested	The dryer could generate a temperature of 36 °C and 22 °C higher under no-load and loaded conditions, respectively, compared to ambient temperatures
8	Islam et al. [27]	2019	Bangladesh	Apple, banana, guava and pineapple	NCSD with three ventilation arrangements was constructed and tested	Maximum moisture removal in the simple ventilation drying chamber was 58.9% which is 14.4% and 25.6% higher than chimney and attic types, respectively

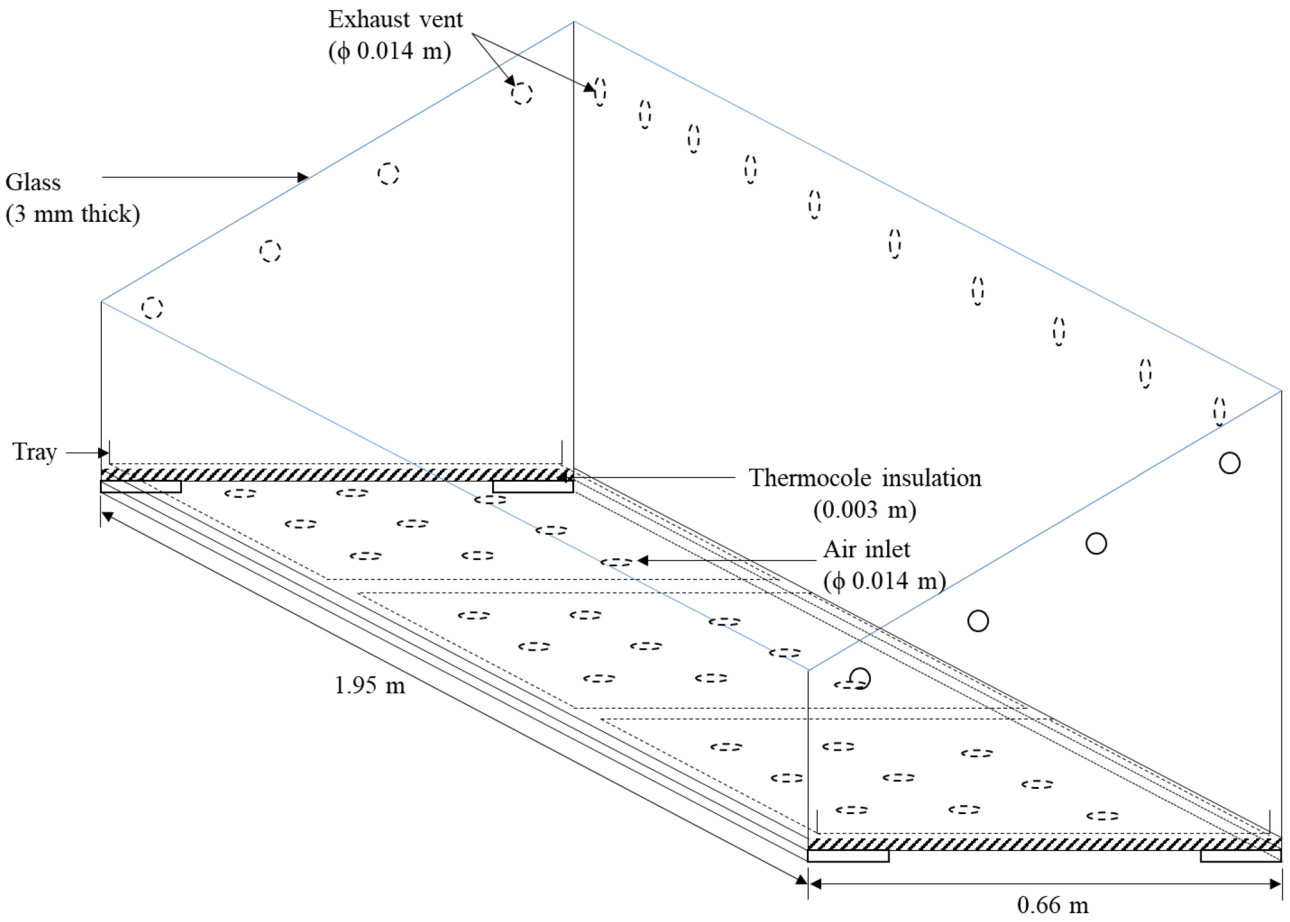


Fig. 6 Schematic view of solar cabinet dryer [28]

Singh et al. [32] designed, developed and tested a domestic NCSD (Fig. 9) under no-load and loaded conditions. Thermal efficiency of the NCSD ranged from 5.16 to 28.05% and 19.3 to 25.6% for batch mode and semi-continuous modes, respectively. Rawat et al. [33] fabricated an NCSD (Fig. 10) with a quadrilateral-shaped cabinet. Moisture content of the chili sample was reduced from 74.4 to 4.4% (w.b.) in 8 h. Alonge and Omoniwa [34] modified a direct NCSD for easier operations with different tilt angles. Dryer was tested under no-load conditions and also for the drying of 1 kg of cassava chips. Maximum difference of 20 °C was observed between the drying chamber and ambient temperature under no-load conditions. Dryer took 32 h in comparison to 38 h under OSD. Alonge and Adeboye [35] evaluated the drying rate of

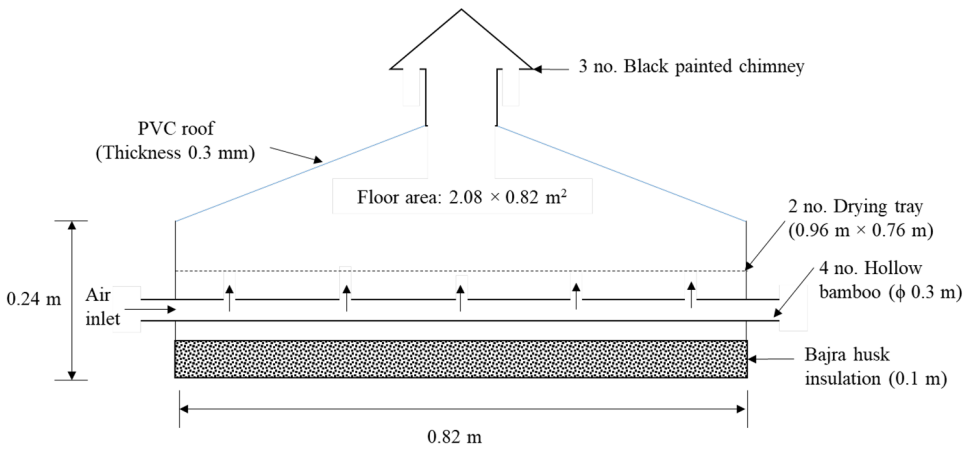


Fig. 7 Direct type low-cost solar dryer [29]

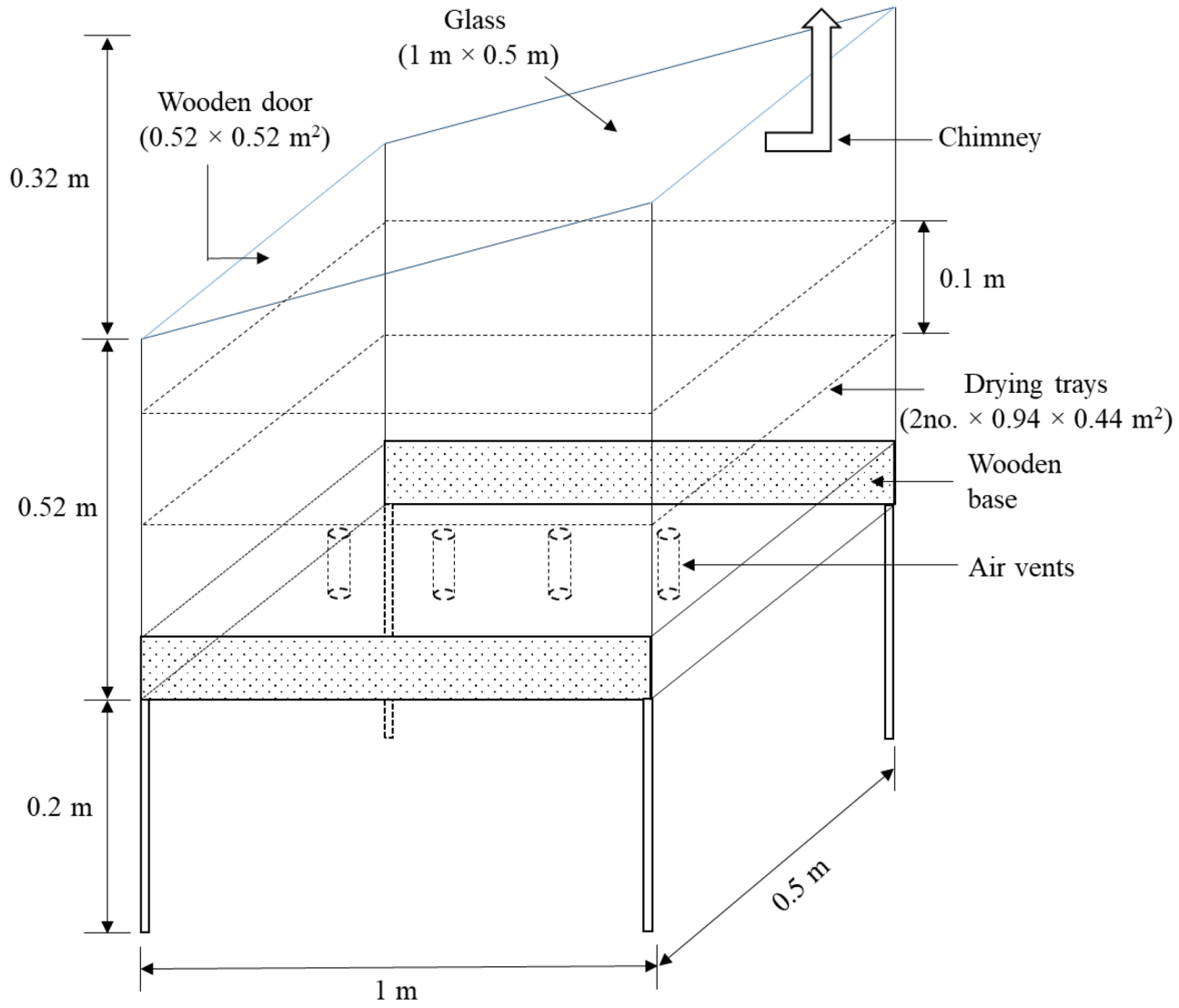


Fig. 8 Domestic solar grain dryer [31]

pepper (Capsicum annum L.), okra (Abelmoschus esculentus L.) and vegetables (Amaranthus hybridus L.) for the respective moisture content of 78.9–24%, 92–20% and 90–20% (w.b.) in a direct and indirect NCSD and also compared with OSD. Drying rates for all three products were observed to be higher in direct type NCSD (3.94 g/h, 17.65 g/h and 13.33 g/h) as compared to the indirect NCSD (2.55 g/h, 15.79 g/h and 11.11 g/h) and OSD (2.17 g/h, 14.29 g/h and 8.33 g/h). Tefera et al. [36] compared the performances of box and pyramid-type solar dryers (Fig. 11) to dry potato slices. A little reduction of about 2–3 h in the drying time with uneven drying in different trays was reported compared to OSD.

Eke [37] developed and tested a direct NCSD for drying tomatoes, carrots and okra in rural Nigerian climates. Drying efficiency was reported to be 21.18–24.95% for NCSD and 10.59–15.19% under OSD. Raju et al. [38] designed and tested an efficient NCSD with an exhaust heat utilization facility to dry 3 kg of potato, bitter gourd and green chili, as shown in Fig. 12. Highest moisture evaporation under NCSD was observed to be 84.33% (w.b.) for bitter gourd followed by 81.66% (w.b.) for potato and 41.5% (w.b.) for chili. A water heating system was installed at the exhaust vent of the drying chamber so that heat going waste from the drying chamber could be used. Eke [39] also compared the performance of the direct NCSD with solar collectors having different absorbing materials (metal, wood, cement and mud) for the drying of sliced tomatoes. The drying chamber and solar collector sizes were 1.53×0.61 m^2 and 0.92×0.61 m^2, respectively. For OSD, wood, cement, mud, and metal type solar collector, the drying efficiencies were 7.38, 19.56, 20.25, 20.91, and 27.24%, respectively. Navale et al. [12]

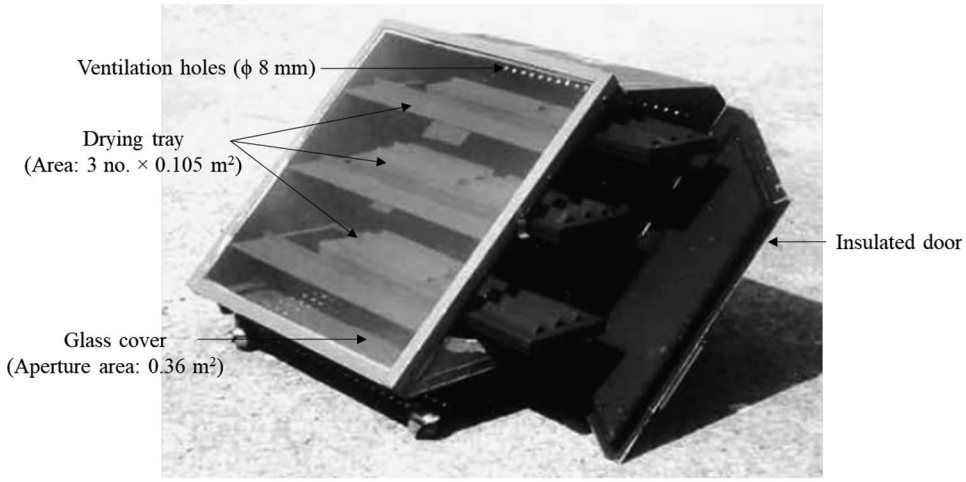


Fig. 9 Multi-shelf domestic solar dryer [32]

designed and developed a cabinet-type NCSD (Fig. 13) and tested it for fenugreek leaves drying. The NCSD maintained a maximum temperature difference of about 20–22 °C as compared to the ambient. Average drying efficiency for the NCSD was 34.50%, 13.98% higher than OSD. Borah et al. [40] experimented with whole and sliced turmeric samples to evaluate drying kinetics using an SCD shown in Fig. 14. SCD reduced the drying time from 25.5 h in OSD to 12 h. SCD was also tested by Subedi and Bhattarai [41] to dry ginger. Overall efficiency was observed to be reduced with an increase in slice thickness and it was maximum (64.36%) with the least slice thickness (3-5 mm). Chaudhari and Bhavsar [42] developed a box-type solar dryer cum cooker (SD-C) as shown in Fig. 15 and tested it under the no-load and loaded condition for chili drying. The moisture content of 3 kg of green chili was reduced from 80.58 to 6.89% (w.b.) in 23 h and 18 h using only solar energy and solar plus electrical energy for the drying operation. The drying efficiency of the SD-C was found to be 19.12 and 19.50% without and with an electrical heating system, respectively. Chaudhari et al. [43] tested a hybrid NCSD for the drying of ginger that took 9 h to reduce the moisture content from 79.31 to 6.73% (w.b.), whereas the reduction was just up to 66.38% (w.b.) under OSD.

Haque et al. [44] developed a portable and durable domestic NCSD (Fig. 16) on a 'do it yourself' model specifically for remote areas and tested for the drying of bitter gourd, okra, hirda and raw mango. Collector and exergy efficiencies were 21–38% and 17–44%, respectively. Ugwuoke et al. [45] developed a mixed-mode solar dryer (Fig. 17)

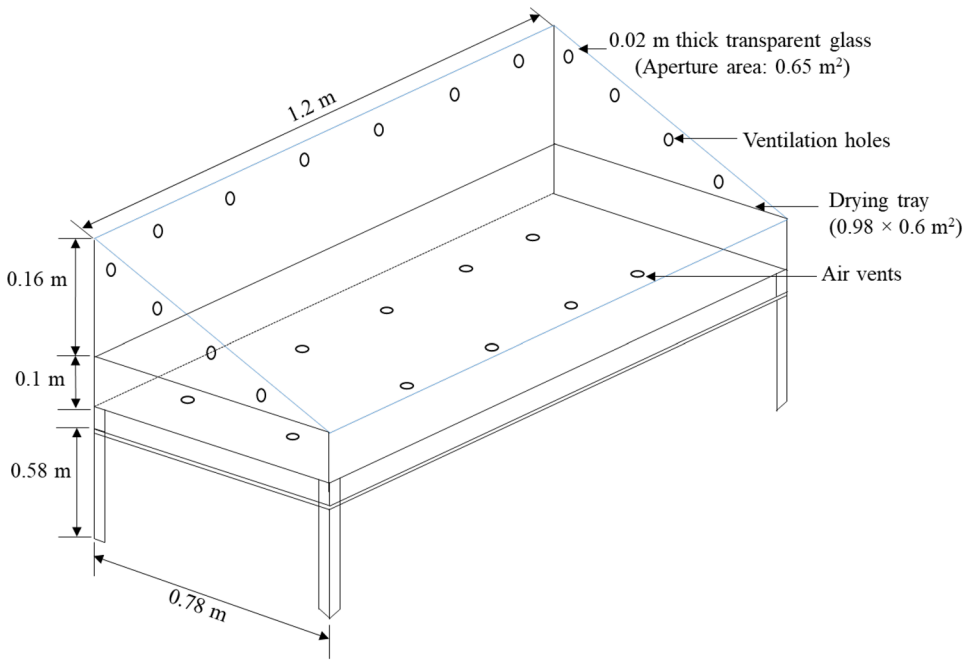
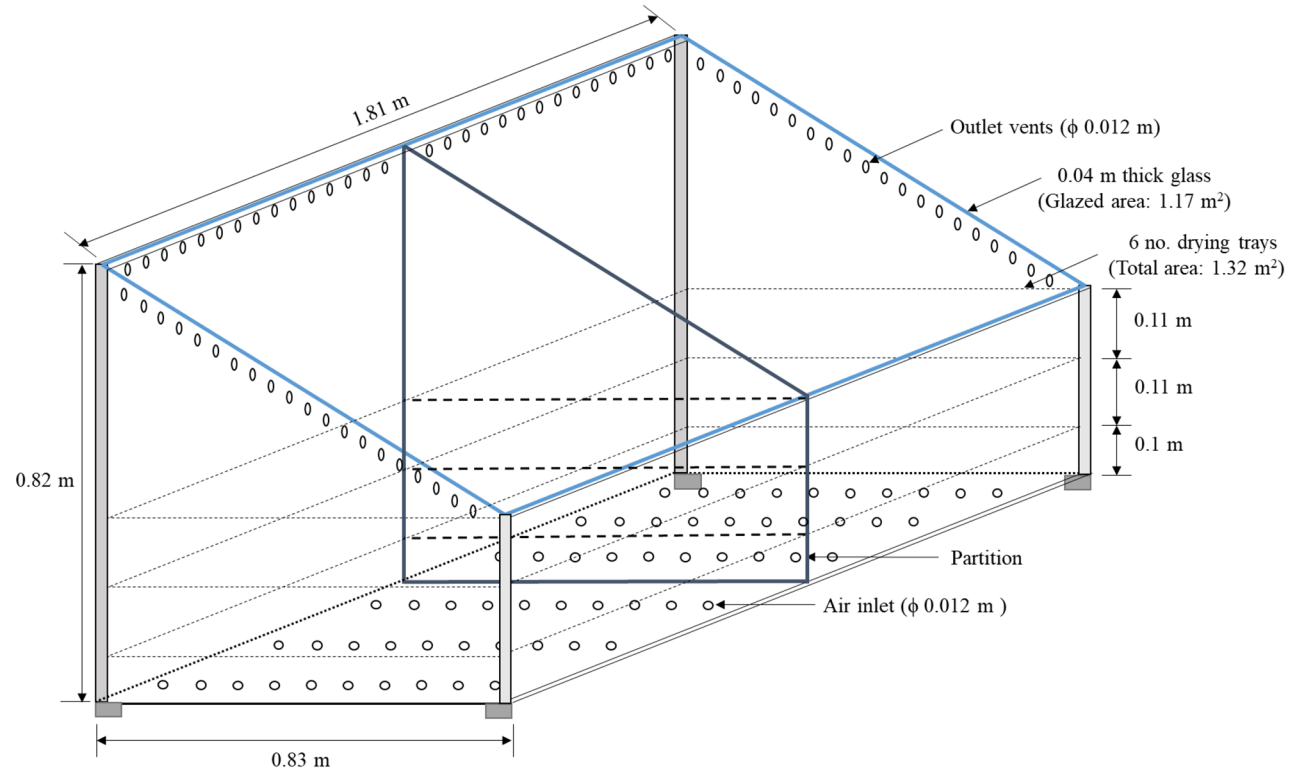
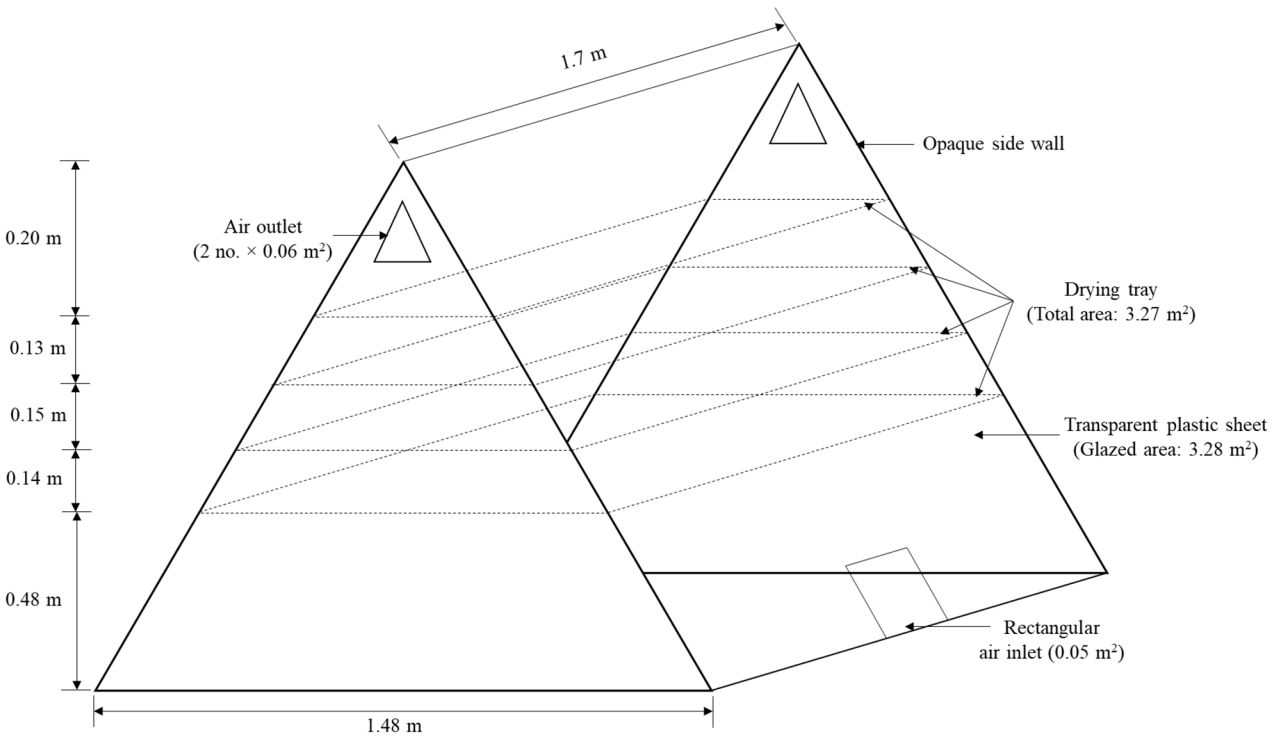


Fig. 10 Natural convection type solar crop dryer [33]



(a)



(b)

Fig. 11 **a** Box type solar dryer, **b** Pyramid type solar dryer [36]

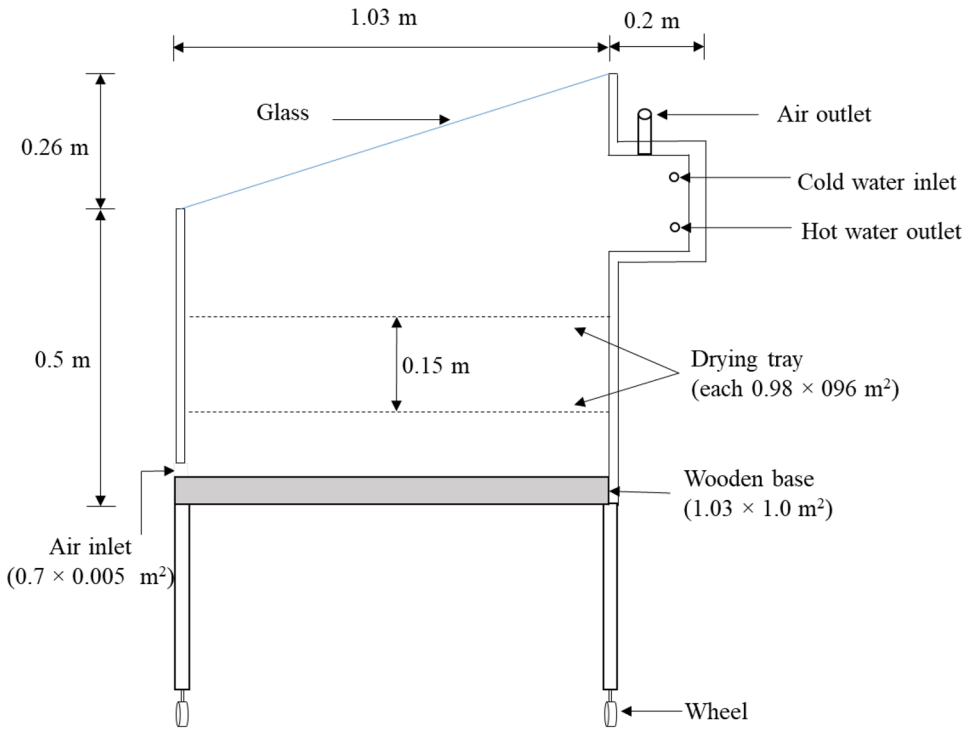


Fig. 12 Schematic of efficient solar dryer [38]

for domestic applications. Temperature inside the drying chamber was 16 °C higher than the ambient temperature. The total moisture loss from pepper, okra and yam slices was 4 g, 10 g and 37 g higher, respectively, compared to OSD. Chavan et al. [46] carried out a computer simulation for SCD and validated the results with experimental observations for drying potato slices. Good agreement was reported in the outcomes of the two analyses. Progress on domestic NCSD for vegetable drying has been summarised in Table 2.

Investigations on Domestic NCSD for Miscellaneous Commodities Drying

Ampratwum and Dorvlo [47] designed and tested a cabinet NCSD (Fig. 18) under no-load conditions at Sultan Qaboos University, Sultanate of Oman. Maximum average temperature for the drying chamber was 84 °C.

Bolaji [48] developed an indirect NCSD with a solar air collector having box-type absorber (Fig. 19). Maximum efficiency was 60.5%, which was 39.5% and 24.5% higher than flat and finned solar collectors. Mwithiga and Kigo [49] designed and tested a box-type NCSD (Fig. 20) for the drying of parchment coffee by using four sun-tracking methods. NCSD reduced the moisture content from 54.8 to 13% (w.b.). However, no significant difference was observed in drying time using various sun-tracking methods. Kumar et al. [50] designed and tested a solar cooker cum dryer (SCCD) for domestic applications, as shown in Fig. 21. Maximum temperatures at the absorbing plate under no-load and water-loaded conditions were 140 °C and 98.6 °C, respectively. It was reported that tracking the sun had reduced the performance of SCCD.

Saleh and Badran [51] developed and tested a domestic NCSD (size: 70 cm × 60 cm × 29 cm) under no-load and loaded conditions for drying Jew's mallow in fixed and tracking mode. NCSD had a provision to accommodate three drying trays and get tilted for higher energy collections. Stagnation temperature and average heat loss coefficient were 80 °C and 9.6 W/m^{2}K, respectively. Abdullahi et al. [52] designed and tested an adjustable and collapsible solar dryer (Fig. 22) to dry groundnuts.

Ozuomba et al. [53] developed a direct absorption solar dryer having a surface area of 0.39 m^2. A 2.5 cm thick softwood was used to fabricate a rectangular drying chamber. Drying bed was created by covering the base with a foam layer, aluminum sheet, and black leather sheet. A sliding door opening (size: 10 cm × 6 cm) was created as an inlet for fresh air and a 5 cm diameter opening was used for the outlet. Mustapha et al. [54] constructed five solar dryers with different materials (plastic, mosquito net, glass, aluminum and glass with black pebbles) and tested them for fish drying. Floor and drying tray sizes were kept as 2 ft × 2 ft and 1.7 ft × 1.7 ft, respectively. Akpojaro and Oyeyemi [55] developed a box-type NCSD for maize cobs. Dryer took 2 days instead of 6 days taken in OSD for 5 kg

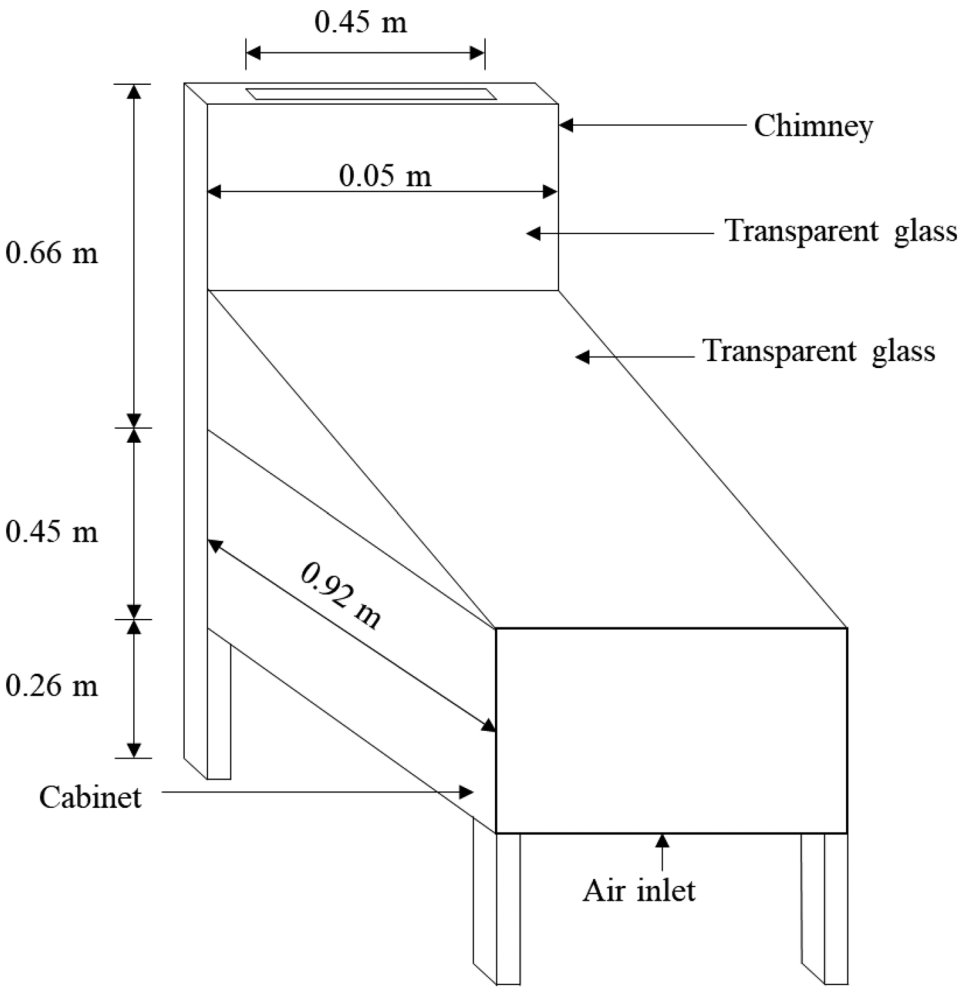


Fig. 13 Cabinet solar dryer [12]

of maize cobs. Chavan et al. [56] analyzed SCD using CFD analysis with two tray modifications to predict the temperature at the top and bottom plates. An increase in temperature and velocity was reported due to suggested modifications. Sandali et al. [57] investigated the effect of the porous medium (gravel, steel wool and stone bed) as sensible heat storage in a direct NCSD (Fig. 23). Drying temperature was increased by 4 °C and the working time was extended by 2 h after sunset with the proposed modification. Mehata et al. [58] tested a mixed-mode solar dryer under no-load conditions for small-scale applications. Jain et al. [59] carried out a CFD simulation for a domestic NCSD. Convective heat transfer coefficient was 2.81 W/m^2 °C, with a significant temperature of about 55 °C around the tray. Reliability of the results was observed in terms of the coefficient of determination, which was reported as 0.98. Sandali et al. [60] analyzed the effect of various heat supply techniques (heat exchange, porous medium and PCM) on a direct NCSD. Investigations on domestic NCSD for the drying of miscellaneous commodities are also summarised in Table 3.

Forced Convection Solar Dryers (FCSD)

The flow rate of heated air through the drying chamber significantly influences the drying rate, drying time, and overall performance of the drying system. Forced convection solar dryers (FCSD) have shown an edge over NCSD in many studies. In this section, various investigations on the use of FCSD for domestic and small-scale farm applications, especially for the drying of fruits, vegetables and some other miscellaneous commodities, have been presented.

Investigations on Domestic FCSD for Fruit Drying

Poonia et al. [61] developed a direct hybrid photovoltaic/thermal (PV/T) FCSD (Fig. 24) for the drying of ber (Zizyphus mauritiana) fruit from an initial moisture content of 80 to 20% (w.b.).

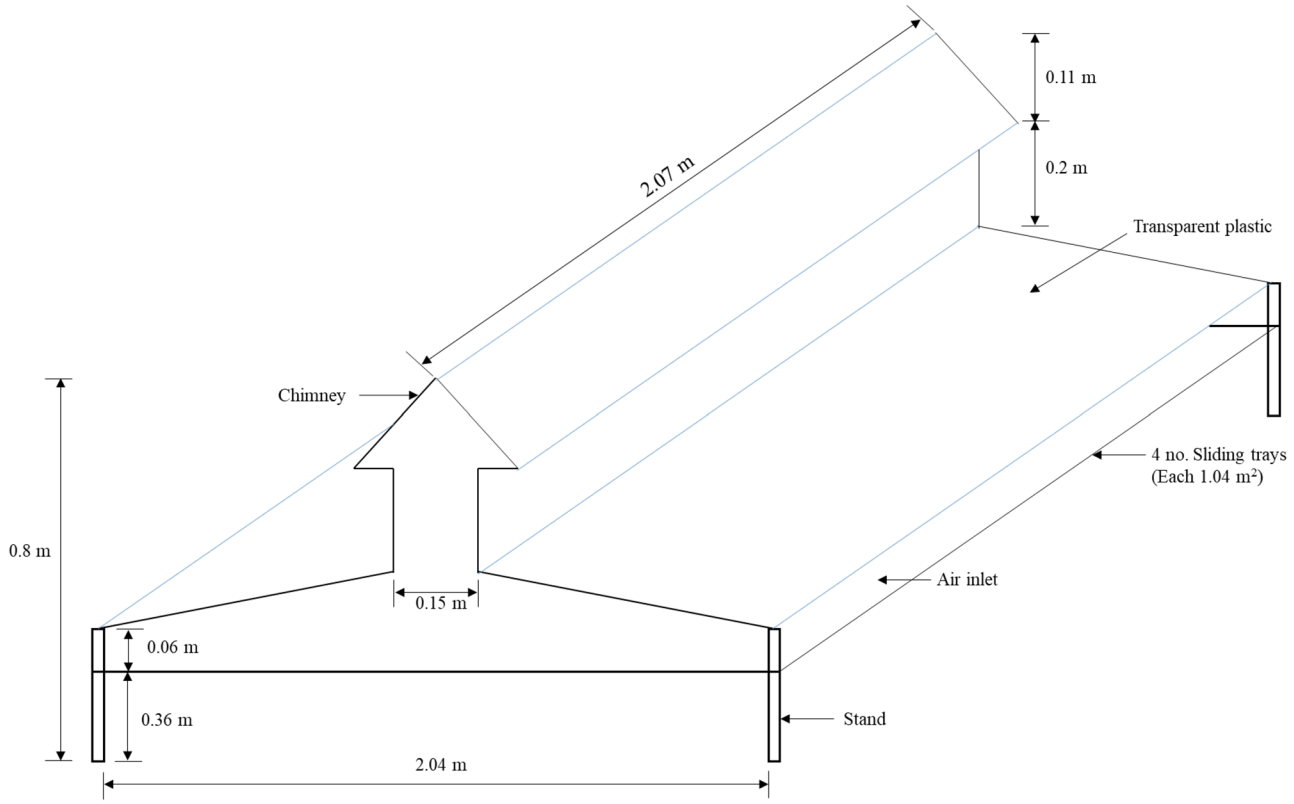


Fig. 14 Schematic of a solar conduction dryer [40]

Poonia et al. [62] compared the performance of a hybrid PV/T solar dryer under natural and forced convection modes to dry Indian jujube (Zizyphus mauritiana). Average thermal efficiency of the dryer was found to be 15.6 and 16.7% for natural and forced convection modes, respectively. Nabnean and Nimnuan [63] used a parabolic-shaped direct FCSD

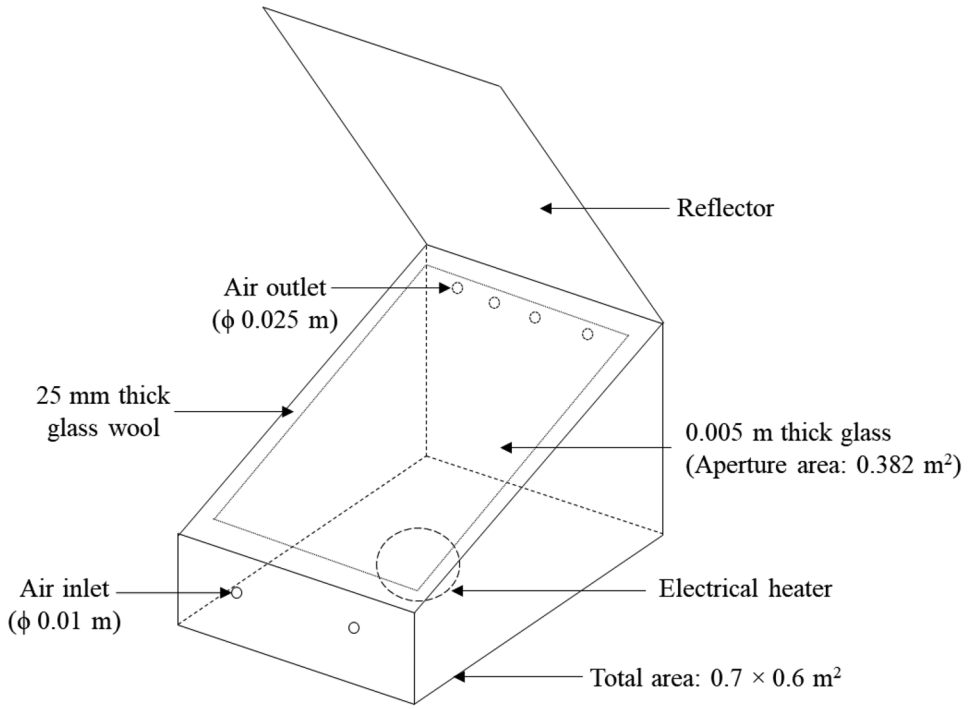


Fig. 15 Hybrid solar dryer cum cooker [42]

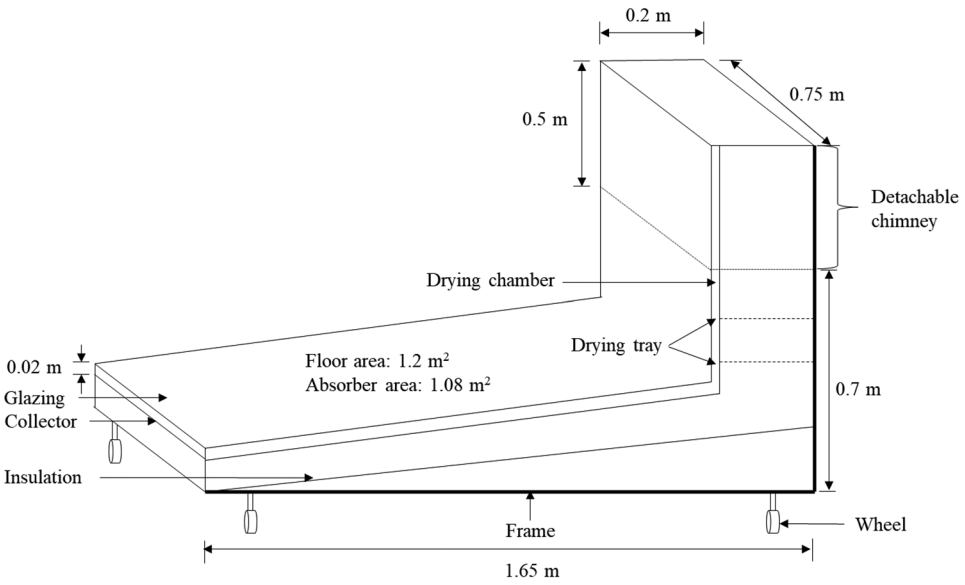


Fig. 16 Schematic of a domestic solar dryer [44]

(Fig. 25) to dry banana slices and compared the results with OSD. Temperature of the drying air varied from 35 to 60 °C. Moisture content was reduced from 72 to 28% (w.b.) in 4 days by using FCSD, whereas it was just reduced up to 40% under OSD in the same drying time. Table 4 summarizes the studies on domestic FCSD for fruit drying.

Investigations on Domestic FCSD for Vegetable Drying

Tiris et al. [64] developed a small-scale indirect FCSD (Fig. 26) with an electrical heater and solar air collector to pre-heat the drying air.. Compared to OSD, a significant reduction in the drying time was reported for drying sultana grapes, green beans, sweet peppers, and chili peppers.

Sreekumar et al. [65] developed and tested a cabinet-type indirect FCSD (Fig. 27). To avoid the direct exposure of drying commodities to solar radiation, the whole system was divided into two chambers called the collection chamber and the drying chamber. Gaikwad et al. [66] developed a briefcase-type portable solar dryer (Fig. 28) to work under both natural and forced convection modes. An arrangement of an electric heater powered through a battery charged by a solar photovoltaic module was made for off-shine operations. Modi et al. [67] fabricated a cabinet-type direct FCSD to dry tomato slices. Moisture content was reduced from 90.48 to 7.73% (w.b.). Mohsen et al. [68] fabricated a simple

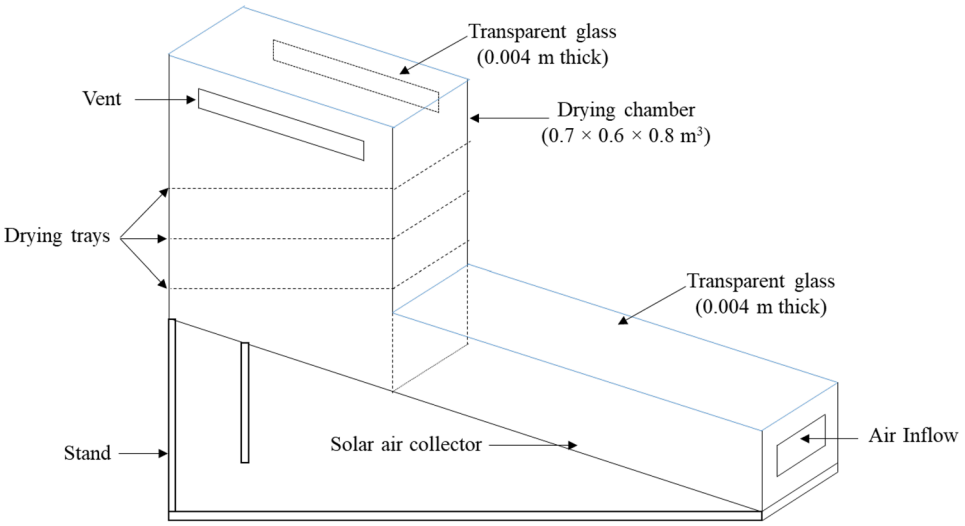


Fig. 17 Mixed-mode domestic solar dryer [45]

Table 2 Summary of progress on domestic NCSD for vegetable drying

Sr. No.	Researcher/s	Year	Country	Drying commodity	Description	Remarks
1	Sharma et al. [28]	1986	India	Peas, grapes and potatoes	Performance of the cabinet-type NCSD was evaluated	NCSD took 3 days for drying with an efficiency of 40%
2	Thanvi and Pande [29]	1987	India	Chilies	Low-cost NCSD was developed and tested	Drying time was reduced by 50% compared to OSD with a maximum efficiency of 37.9%
3	Sharma et al. [30]	1990	India	Turnip, cauliflower and green peas	A cabinet type NCSD was compared with OSD	Drying time was reduced by 4, 12 and 6 h compared to OSD for drying turnip, cauliflower, and green peas, respectively
4	Ezekoye and Enebe [31]	2006	Nigeria	Pepper and groundnuts	NCSD was fabricated for domestic applications	Average collector and drying efficiencies were 10% and 22%, respectively
5	Singh et al. [32]	2006	India	Fenugreek leaves and chilies	NCSD with the variable inclination and covering plate was developed and tested	Thermal efficiency was reduced with time under batch mode. The drying cost for fenugreek leaves was only 60% of an electric dryer
6	Rawat et al. [33]	2009	India	Chili	Evaluated energetics of a direct NCSD	Total energy spent on the NCSD for the entire lifetime was 2744.64 MJ, with 1–2 years of energy payback period
7	Alonge and Omoniwa [34]	2012	Nigeria	Cassava chips	Constructed NCSD for different tilt-angle settings for higher energy collection	Drying time was reduced by 15.7%
8	Alonge and Adeboye [35]	2012	Nigeria	Pepper, okra and vegetables (Amaranthus hybridus L.)	Drying performances of direct and indirect NCSD were evaluated	Direct NCSD took 33, 51 and 30 h for pepper, okra and vegetables drying, which were 35, 10 and 16% lower than indirect NCSD
9	Tefera et al. [36]	2013	Ethiopia	Potato slices	Performances of the box and pyramid-type solar dryers were compared	The cost and weight of the pyramid-type dryer were 72.59% and 73.33% less than the box-type dryer, respectively
10	Eke [37]	2013	Nigeria	Tomato, carrot and okra	Fabricated and tested a small-scale NCSD	The drying efficiencies of NCSD were found to vary from 21.18 to 24.95%
11	Raju et al. [38]	2013	India	Potato, bitter gourd and green chili	NCSD with a provision for exhaust heat utilization was tested and compared with OSD	NCSD reduced the moisture by 81.66, 84.33 and 41.5% (w.b.) as compared to 69.33, 73.66 and 19% (w.b.) under OSD for potato, bitter gourd and chili, respectively
12	Eke [39]	2014	Nigeria	Sliced tomatoes	Direct NCSD with wood, cement, mud, and a metal solar collector was compared with OSD	Drying time was reduced by 56.75–65.76%
13	Navale et al. [12]	2015	India	Fenugreek leaves	Compared the performance of NCSD having variable width chimney with OSD	Using NCSD, drying time was reduced by 42.8% and the energy utilization rate was improved by 75%
14	Borah et al. [40]	2015	India	Turmeric rhizomes	Drying kinetics was evaluated for whole and sliced turmeric samples under SCD	Drying time was reduced by 52.9%, with an overall efficiency of 55%. Sliced samples gave better drying results and the page model was found the most suitable

Table 2 (continued)

Sr. No.	Researcher/s	Year	Country	Drying commodity	Description	Remarks
15	Subedi and Bhattarai [41]	2017	Nepal	Ginger	SCD was tested for different slice thicknesses	Maximum efficiency was 64.36% for 3–5 mm slice thickness
16	Chaudhari and Bhavsar [42]	2017	India	Green chili	SD-C with an electric heater as a modification was tested	Drying time was reduced by 21.7%, with a slight improvement in drying efficiency
17	Chaudhari et al. [43]	2018	India	Ginger	The performance of a hybrid NCSD having a heating element was evaluated	Drying efficiency was 19.71% and the estimated daily benefit was $0.89
18	Haque et al. [44]	2018	India	Bitter gourd, okra, hirda and raw mango	A low cost durable domestic NCSD was tested for rural applications	The maximum temperature difference between the drying chamber and the ambient was around 15 °C and the drying efficiency was 10.73%
19	Ugwuoke et al. [45]	2019	Nigeria	Pepper, okra and yam slices	A mixed-mode NCSD was fabricated and tested for domestic applications	A maximum temperature of 55 °C was observed inside the drying chamber when the ambient temperature was 39 °C
20	Chavan et al. [46]	2020	India	Potato slice	Experimental and simulation analyses were conducted for SCD	Energy transferred to the sample by conduction, convection, and radiation was 34.96, 0.45, and 9.99 W, respectively

direct FCSD and tested the same for tomatoes drying. The performance of the FCSD was compared with OSD for different slice thicknesses (1, 1.5 and 2 cm) and slice layers (single, double and triple). A maximum decrease of 56% in the drying time was observed in single-layer drying with a slice thickness of 1 cm of the product. Moghimi et al. [69] designed and developed an indirect FCSD (Fig. 29) and tested the same for the drying of tomato slices (thickness of 7 mm) for small and household applications. The summary of domestic FCSD for vegetable drying is presented in Table 5.

Investigations on Domestic FCSD for Miscellaneous Commodities Drying

Fagunwa et al. [70] developed and tested a mixed-mode solar dryer (Fig. 30) with a thermal energy storage unit for cocoa beans under the Nigerian climate. Society for Energy, Environment and Development (SEED) developed a direct solar cabinet dryer (Fig. 31) in 1996 to dry Gum Karaya. Dryer reduced the drying time from 10–15 days in OSD to 2–3 days [71].

Seveda and Jhajharia [72] developed and tested a cabinet-type direct FCSD (Fig. 32) to dry large cardamom (Amomum subulatum). Moisture content was reduced from 75.6 to 10.1% (w.b.) in 24 h, whereas it took 48 h for the same in the case of OSD. Jangsawang [73] developed a cabinet FCSD (Fig. 33) with a greenhouse chamber attached at the top separated by a black absorbing aluminum sheet having the arrangement to force the heated air from the head to the drying chamber. Temperature variation of 1–3 °C was observed among the 5 drying trays with a maximum average temperature of 48.63 °C at the top two trays. A higher percentage of weight loss was reported with the FCSD than OSD. Alonge and Obayopo [74] designed a direct type of FCSD to dry fish (Fig. 34) and tested numerically using ANSYS FLUENT and experimentally for optimum dimensions at various airflow rates under no-load and loaded conditions. Maximum temperature and collection efficiency were observed in natural convection mode. Nimnuan and Nabnean [75] tested a parabolic shaped direct FCSD for the drying of Galangal (Alpinia galangal (Linn.) Swartz.). Temperature inside the drying chamber was in the range of 30–55 °C during the operation. Moisture content was reduced from 89 to 12% (w.b.), saving 43% of drying time compared to OSD. Safri et al. [76] reported the temperature variations in a portable greenhouse dryer with four air collectors (flat plate with insulator, V-groove with insulator, flat plate and V-groove with insulator, and flat plate and V-groove without insulator). Collector with flat plate and V-groove

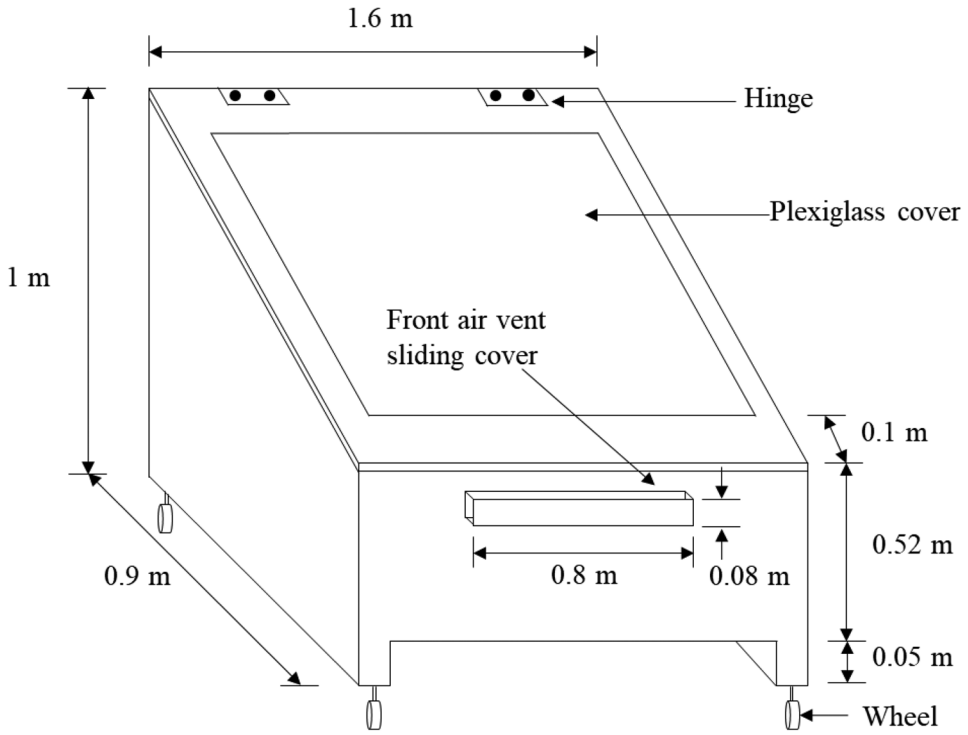


Fig. 18 Solar cabinet dryer [47]

without insulator showed the highest efficiency (66.34%) and was recommended for drying purposes due to higher temperatures inside the drying chamber. Investigations on domestic FCSD for miscellaneous commodities drying are summarised in Table 6.

Economic Analysis of Various Domestic Solar Dryers

Traditional OSD technique has several drawbacks in terms of product quality and drying performance, but there is a

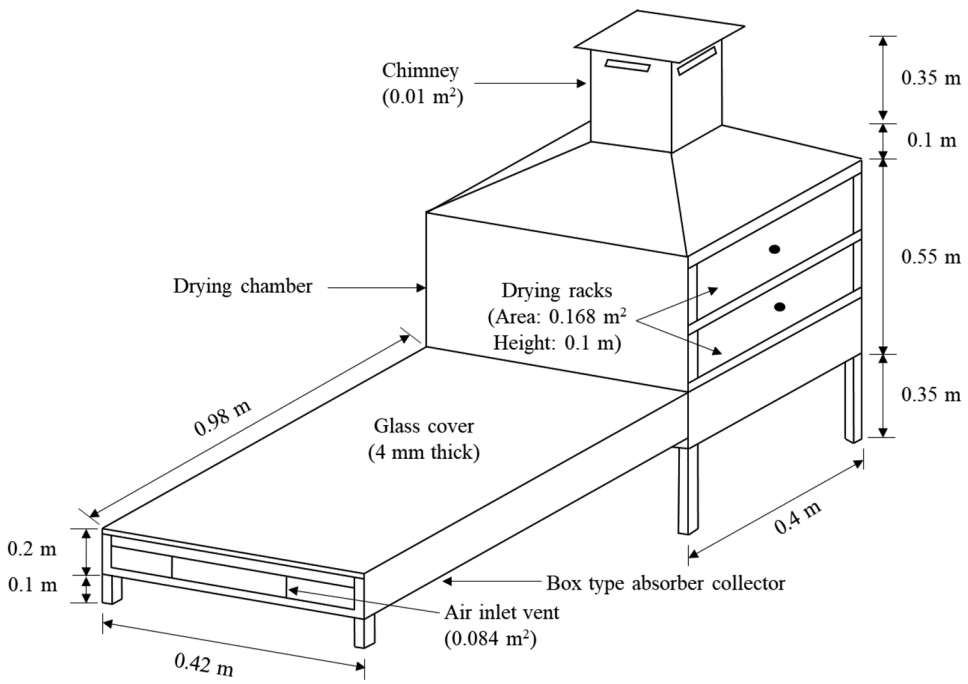


Fig. 19 Solar crop dryer with box type absorber collector [48]

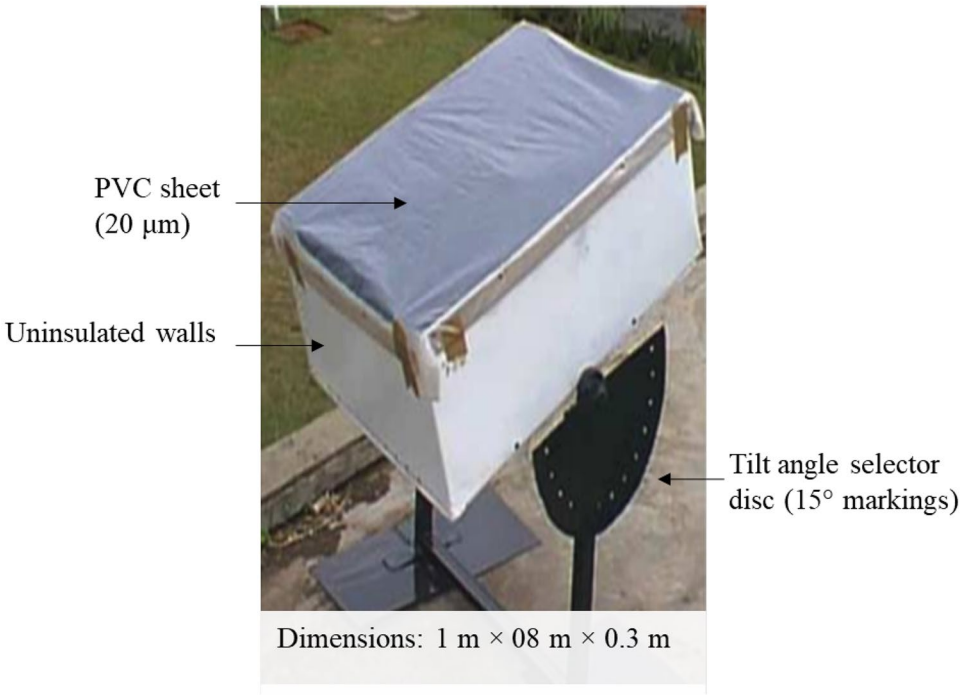


Fig. 20 Photograph of NCSD with partial tracking [49]

major factor responsible for its popularity, i.e., the cost involved in the process is zero. This is more than enough to understand the role of economic analysis for various solar drying technologies. Comprehensive analysis gives the overall feasibility in terms of development and utilization. The comparison among profits that can be achieved by utilizing different solar drying technologies and traditional sun drying can directly influence people more than

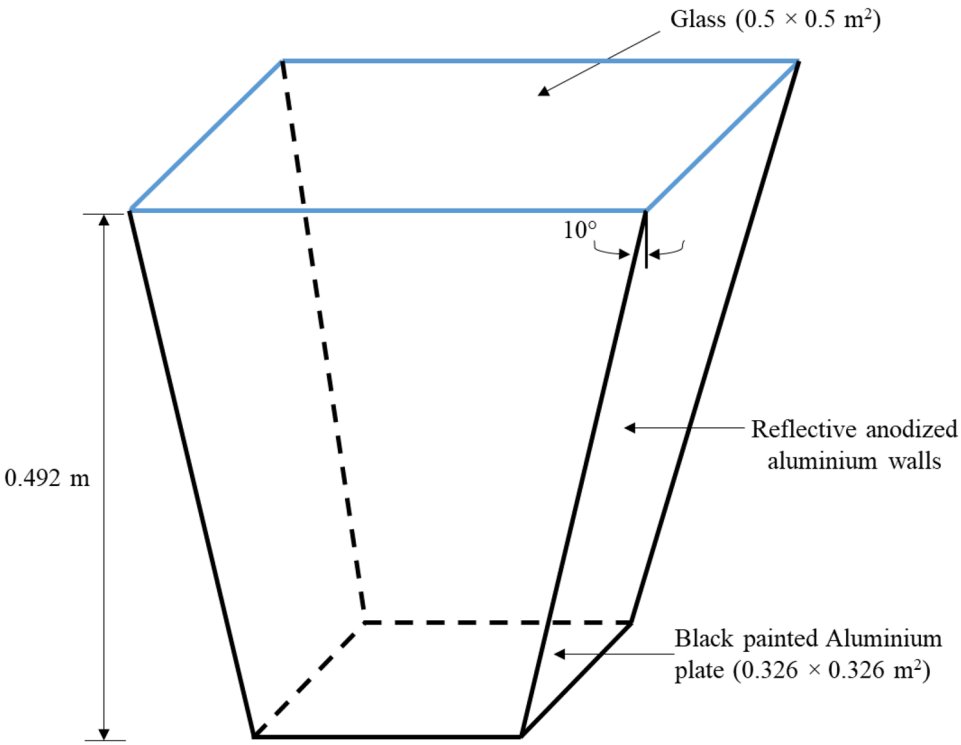


Fig. 21 Truncated pyramid type SCCD [50]

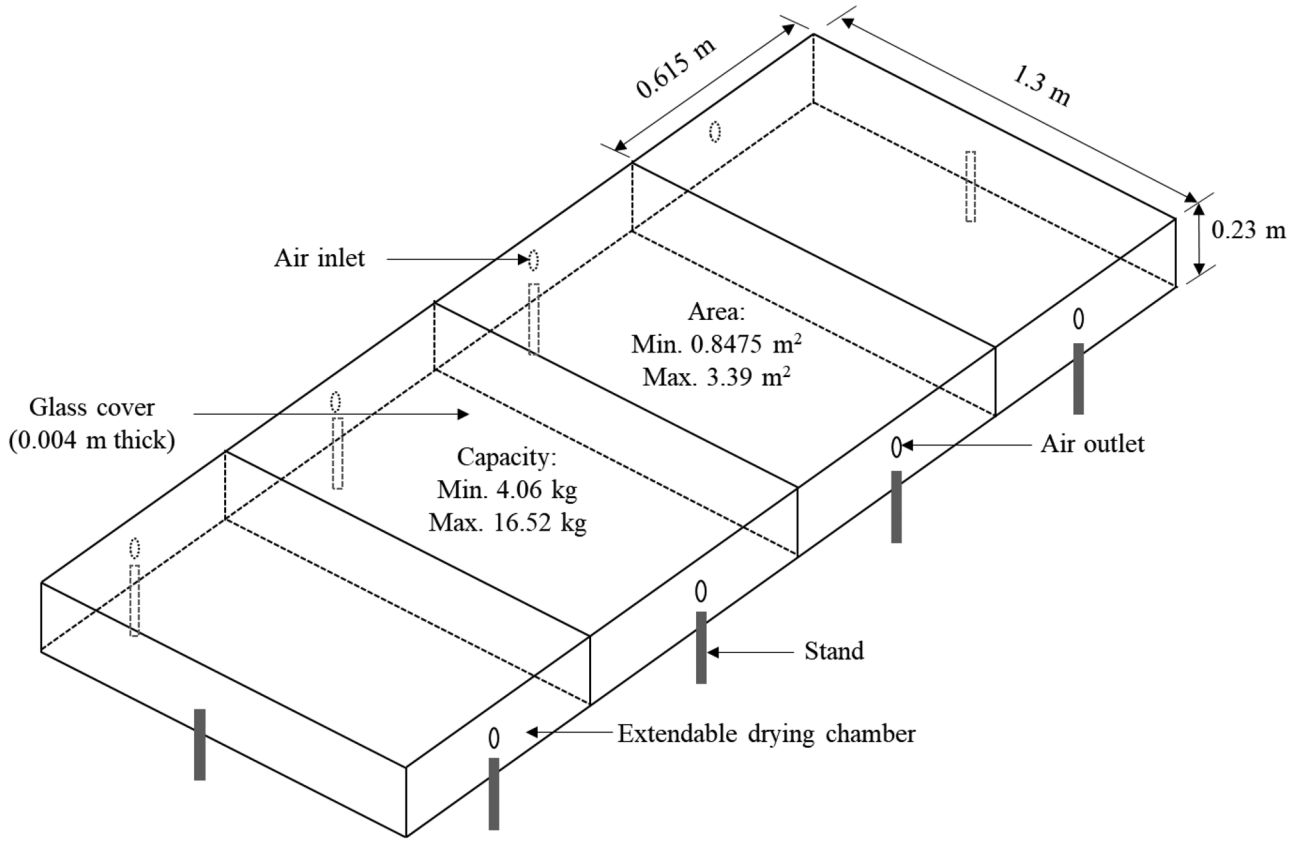


Fig. 22 Adjustable and collapsible NCSD [52]

anything else. Various economical parameters such as total cost of the system, benefit to cost ratio, net present worth and payback time period were being evaluated by different researchers. Economic investigations carried out by different researchers for domestic solar dryers are summarised in Table 7.

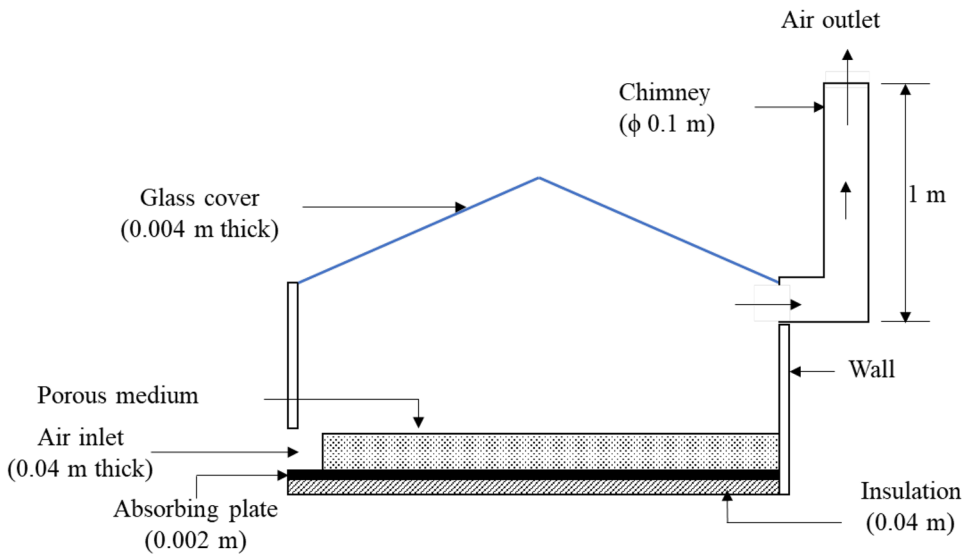


Fig. 23 Schematic of direct solar dryer with porous medium [57]

Table 3 Summary of domestic NCSD for miscellaneous commodities drying

Sr. No.	Researcher/s	Year	Country	Drying commodity	Description	Remarks
1	Ampratwum and Dorvlo [47]	1998	Sultanate of Oman	No-load	Tested a cabinet NCSD for temperature range and energy collection	Maximum heat collection was 0.93 kW/m^2
2	Bolaji [48]	2005	Nigeria	No-load	Indirect NCSD with a solar collector having box type absorber was tested	Maximum temperature inside the drying chamber was 41.22% higher than the ambient temperature (33.5 °C)
3	Mwithiga and Kigo [49]	2006	Kenya	Parchment coffee	Box type NCSD with varying tilt angles for better energy collection was tested	Dryer took 2–3 days compared to 5–7 days in OSD. The tracking system was not recommended
4	Kumar et al. [50]	2008	India	No-load, with water	SCCD was tested for domestic and farm applications	With and without sun tracking, the maximum stagnation temperature was 119 and 112 °C, respectively
5	Saleh and Badran [51]	2009	Jordan	Jew's mallow	Tested a multi-shelf domestic NCSD under fixed and sun tracking mode	Thermal efficiency was improved by 3.34% and specific energy consumption was reduced by 47.23% under tracking mode
6	Abdullahi et al. [52]	2013	Nigeria	Groundnuts	A new extendable design of box-type solar dryer was presented and tested	The minimum and maximum extendable drying areas were 0.8475 and 3.39 m^2 with respective capacities of 4.06 and 16.52 kg
7	Ozuomba et al. [53]	2013	Nigeria	No-load	Temperature range of a direct NCSD was observed	A maximum difference of 26 °C was observed between the dryer and ambient temperature
8	Mustapha et al. [54]	2014	Nigeria	Fish	Economics of the five solar dryers made up of five different materials were compared	Net income varied from $10.91 to 18.18 for various solar dryers
9	Akpojaro and Oyeyemi [55]	2015	Nigeria	Maize cobs	A box-type NCSD was developed and tested	Drying time was reduced by 66.6%
10	Chavan et al. [56]	2018	India	–	Conducted CFD simulation of SCD with modified drying plate and absorber	Positioning of an array of plates and grooved absorber plate resulted in improved outlet air velocity and temperature
11	Sandali et al. [57]	2018	Algeria	–	A porous medium as sensible heat storage in a direct NCSD was simulated numerically	Maximum dryer temperature was 51 °C with the steel wool and gravel bed
12	Mehata et al. [58]	2018	India	No-load	A mathematical model was developed for the estimation of the absorber plate outlet temperatures	Average collector outlet temperature was 50 °C with lower relative humidity throughout the day
13	Jain et al. [59]	2019	India	–	CFD simulation was carried out for a domestic NCSD using ANSYS Fluent 14.0 software	The highest values of coefficient of performance and heat utilization factor were 0.46 and 0.69, respectively
14	Sandali et al. [60]	2020	Algeria	–	Different heat supply techniques to a direct NCSD were analyzed	Maximum temperature of 58 °C was observed by using the geothermal water supply heat exchanger

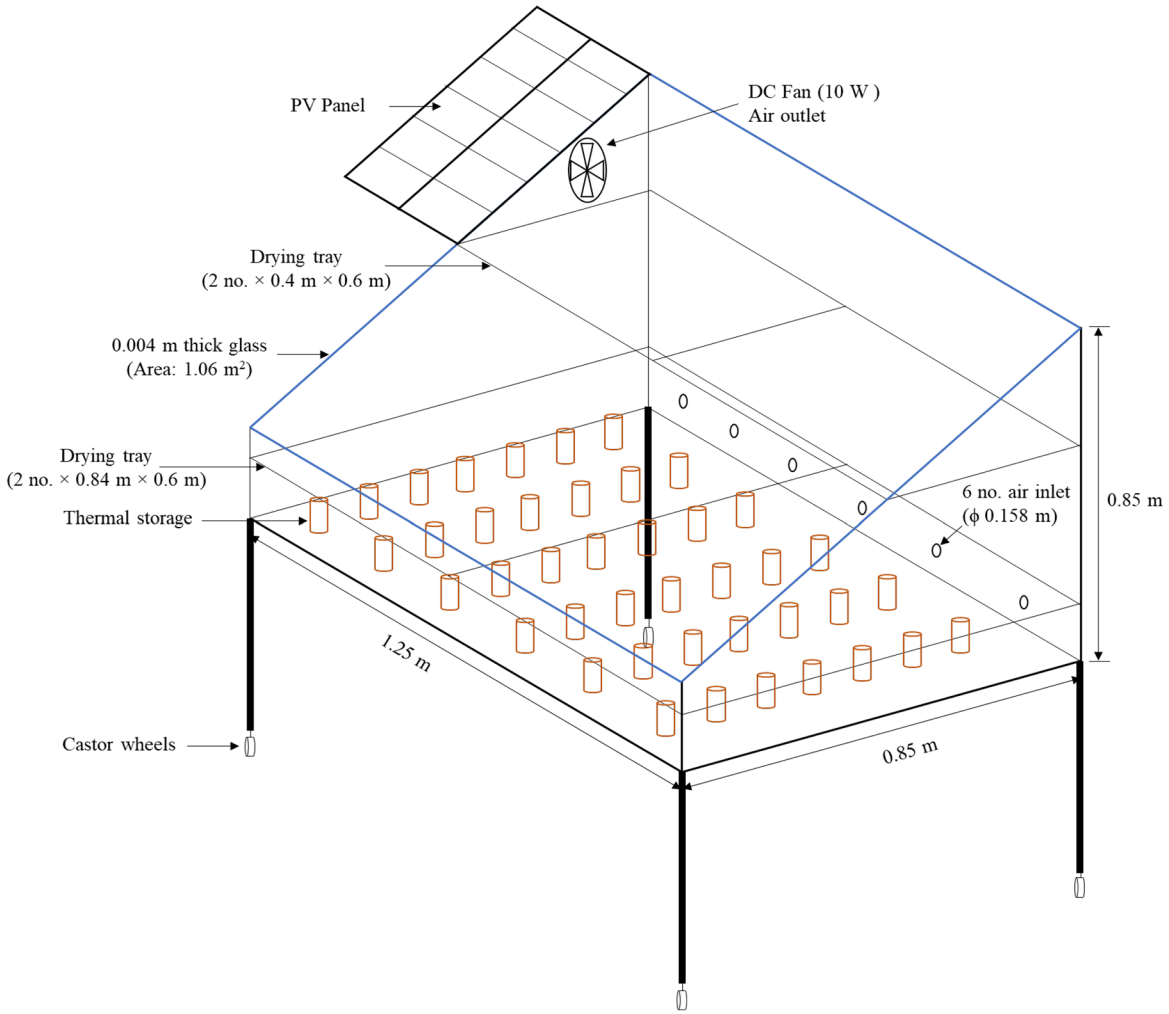


Fig. 24 Schematics of PVT hybrid solar dryer [61]

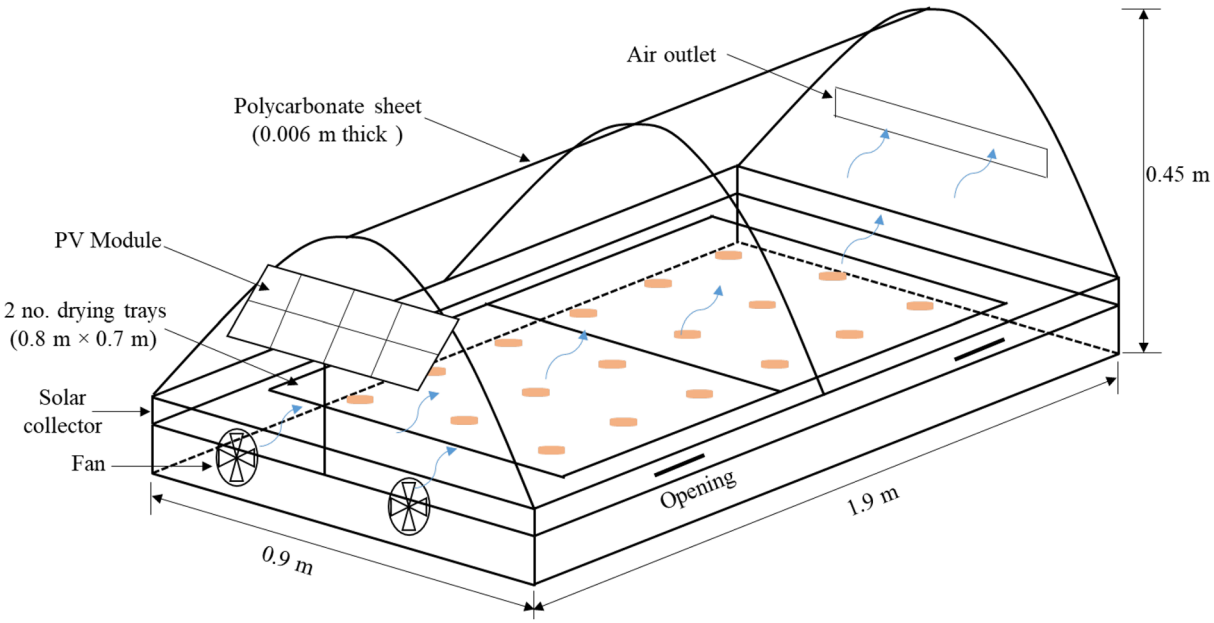


Fig. 25 Parabolic-shaped direct FCSD [63]

Table 4 Summary of domestic FCSD for fruits drying

Sr. No.	Researcher/s	Year	Country	Drying commodity	Description	Remarks
1	Poonia et al. [61]	2018	India	Ber (Zizyphus mauritiana)	Thin layer drying characteristics and economics of hybrid FCSD were presented	The logarithmic model was the most suitable and the efficiency of the FCSD was 16.7% for the drying period of 240 h
2	Poonia et al. [62]	2018	India	Indian jujube (Zizyphus mauritiana)	A hybrid PV/T solar dryer was tested under natural and forced convection modes	Forced mode reduced the drying time by 20%, incrementing average thermal efficiency slightly
3	Nabnean and Nimnuan [63]	2020	Thailand	Banana slices	Parabolic shaped polycarbonate sheet cover solar collector was used as FCSD	Drying time was reduced by 48%

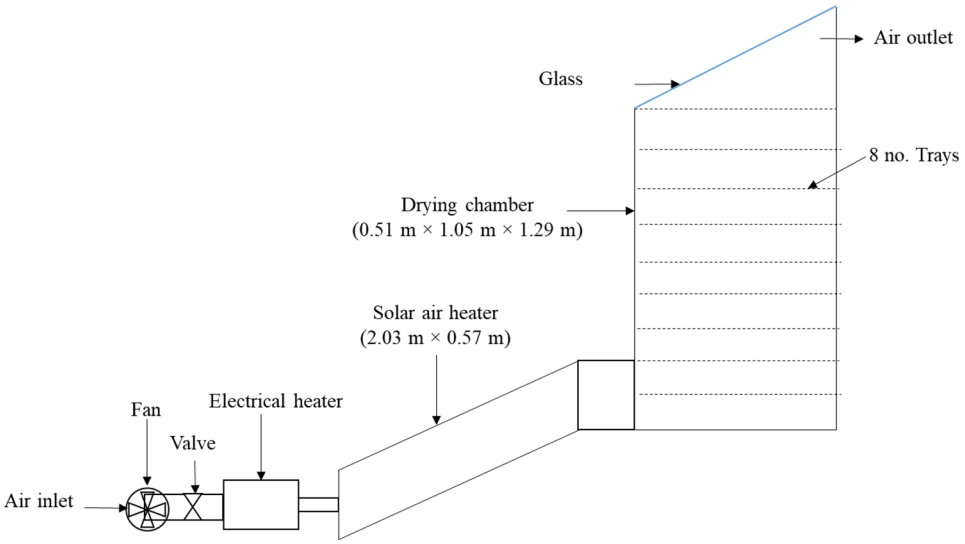


Fig. 26 Schematic of a small scale solar dryer [64]

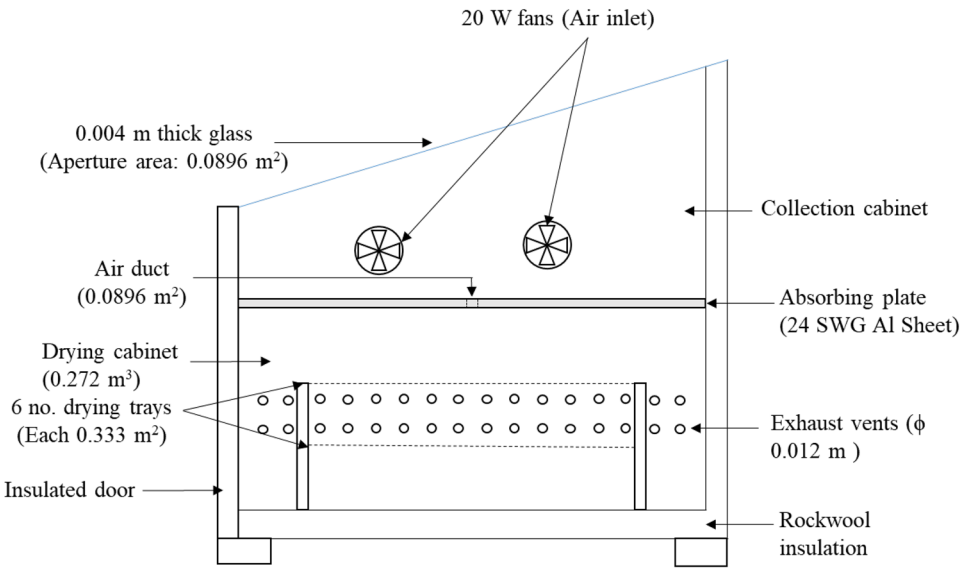


Fig. 27 Indirect solar cabinet-dryer [65]

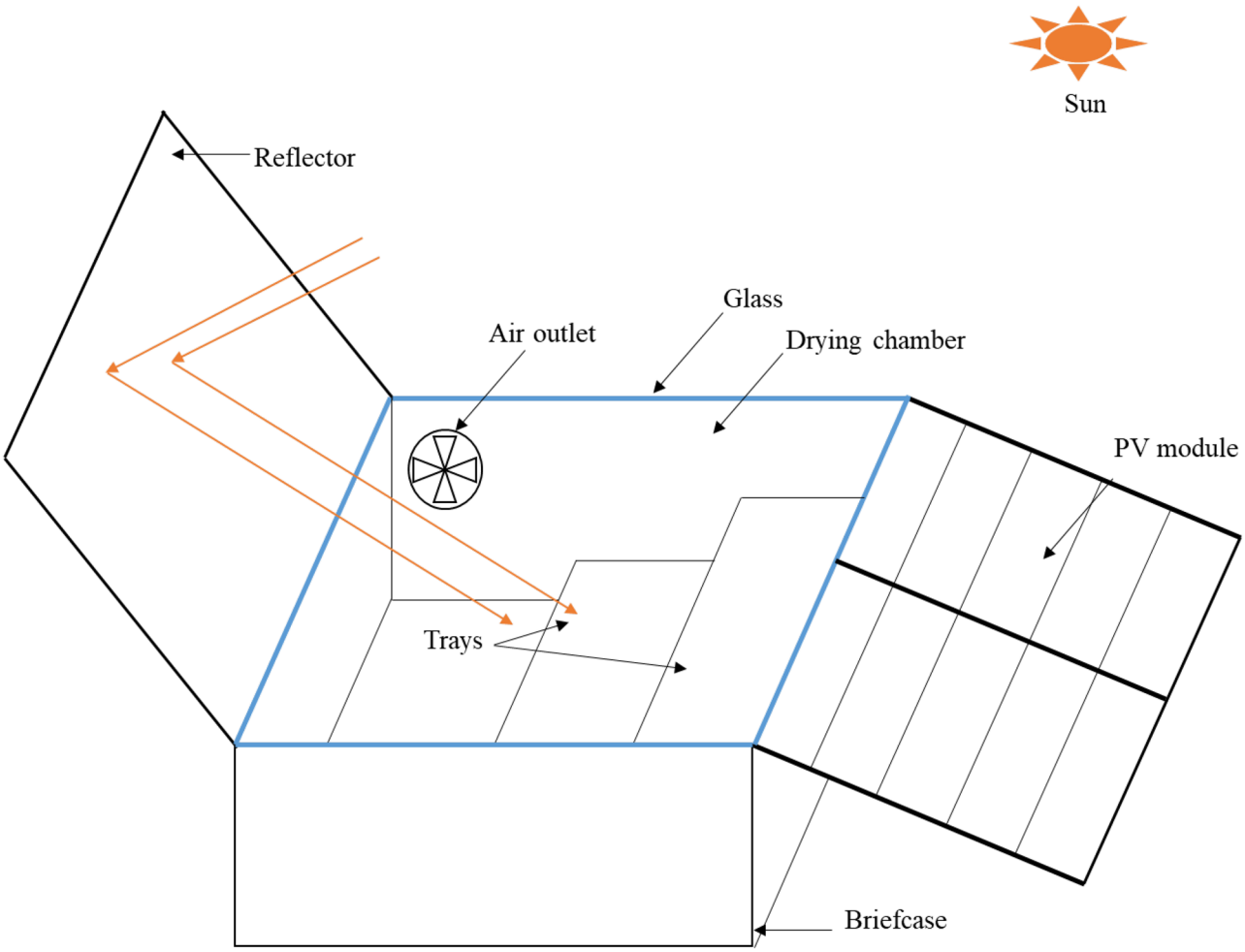


Fig. 28 Briefcase-type portable solar dryer [66]

Summary, Future Scope and Recommendations

The abundance and non-polluting characteristics of solar energy have attracted many researchers for its applications in food drying technologies. Food being a basic necessity, is related to everyone in the world. Domestic solar dryers can give the access to an individual to process food for safe storage and play a significant role in achieving the sustainable development goals suggested by the United Nations, i.e., assuring access to food, eliminating malnutrition, sustainable food production, and affordable clean energy. A summary with some recommendations is given as:

- The simplicity of the design, cost, drying time and drying rates were major factors considered in the development of solar dryers for household applications. It was also observed that many of the studies were confined to very basic designs to maintain simplicity. There were very few investigations on the applicability of indirect and mixed-mode type solar dryers in domestic solar drying.
- Mixed-mode solar dryers must be studied more for domestic uses. Some researchers have shown photovoltaic modules to make an electricity-independent forced convection solar dryer as forced convection solar drying is faster than natural convection solar drying. The thermal efficiency of solar dryers was found to be 50–65%, which can be increased by using various modifications for performance improvement and thermal loss reduction.
- A lack of research was observed on phase change materials as energy storage in domestic solar dryers. Many researchers have recommended using energy storage techniques in large-scale solar dryers. Hence it must be investigated for its suitability in domestic solar dryers.
- Computer simulation tools have shown their significance in design and process optimization. Computer simulation for design optimization is recommended for

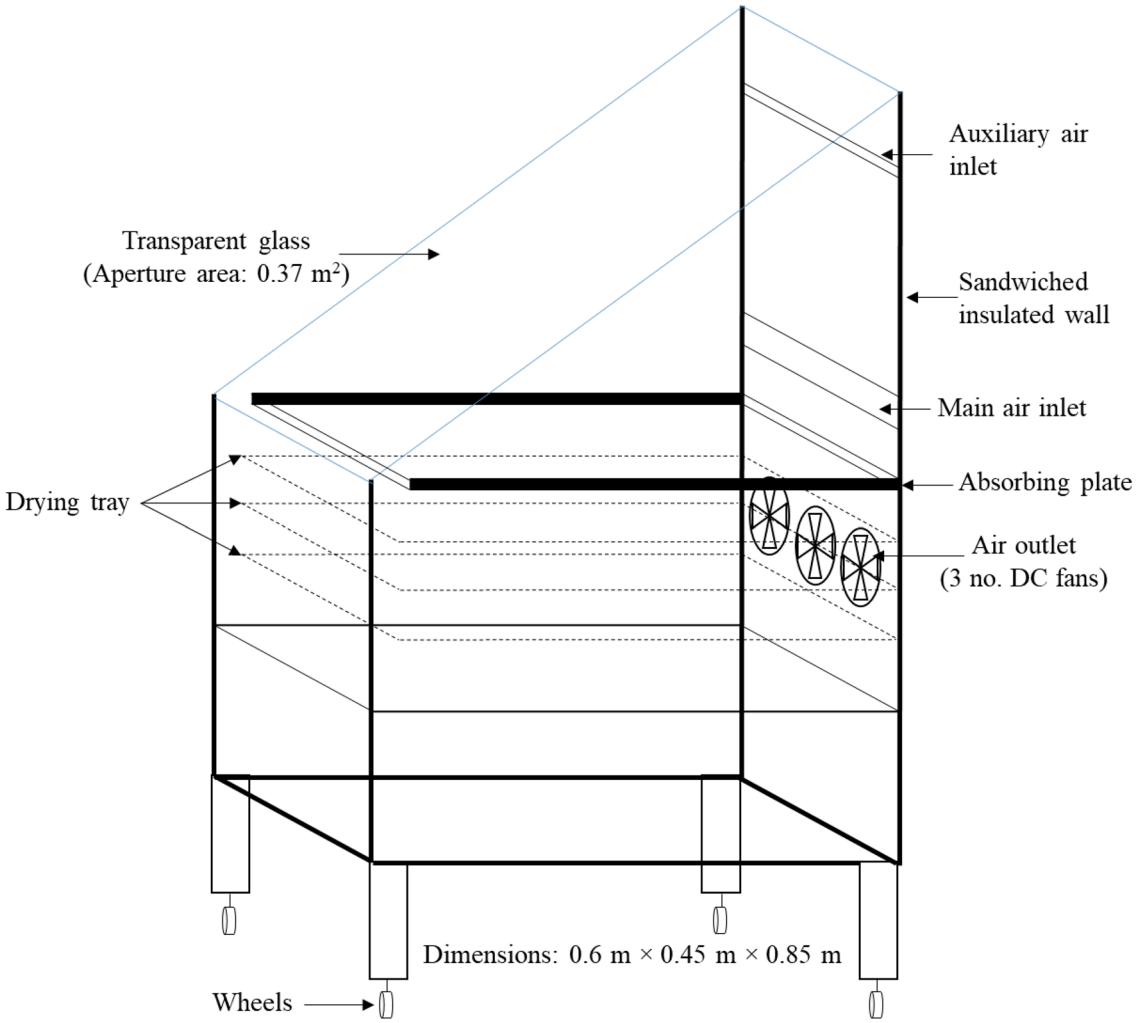


Fig. 29 Small-scale indirect FCSD [69]

Table 5 Summary of domestic FCSD for vegetable drying

Sr. No.	Researcher/s	Year	Country	Drying commodity	Description	Remarks
1	Tiris et al. [64]	1996	Canada	Sultana grapes, green beans, sweet peppers and chili peppers	Indirect FCSD with an electrical heater was tested	Drying time varied from 2 to 5 days for various products
2	Sreekumar et al. [65]	2008	India	Bitter gourd	Investigated a cabinet-type indirect FCSD	Drying time was reduced by 45.45%, with a maximum drying air temperature of 78.1 °C under no-load compared to OSD
3	Gaikwad et al. [66]	2016	India	Potato and tulasi (Ocimum tenuiflorum)	Briefcase-type portable hybrid solar dryer was designed and constructed	Dryer was reported to work under off-shine periods
4	Modi et al. [67]	2017	India	Tomato slices	Drying performance of a direct FCSD was evaluated and compared with OSD	Drying time was reduced by 82.5%, with an average efficiency of 34.87%
5	Mohsen et al. [68]	2019	Egypt	Tomato	The effect of slice thickness and multiple layers on the performance of simple direct FCSD was observed	A single layer of 1 cm thick slices showed the least drying time (50 h), which was 35.8% lower than OSD
6	Moghimi et al. [69]	2021	Iran	Tomato	Indirect FCSD was developed and tested experimentally and numerically	Overall efficiency was 16.4% with a maximum temperature of 65 °C

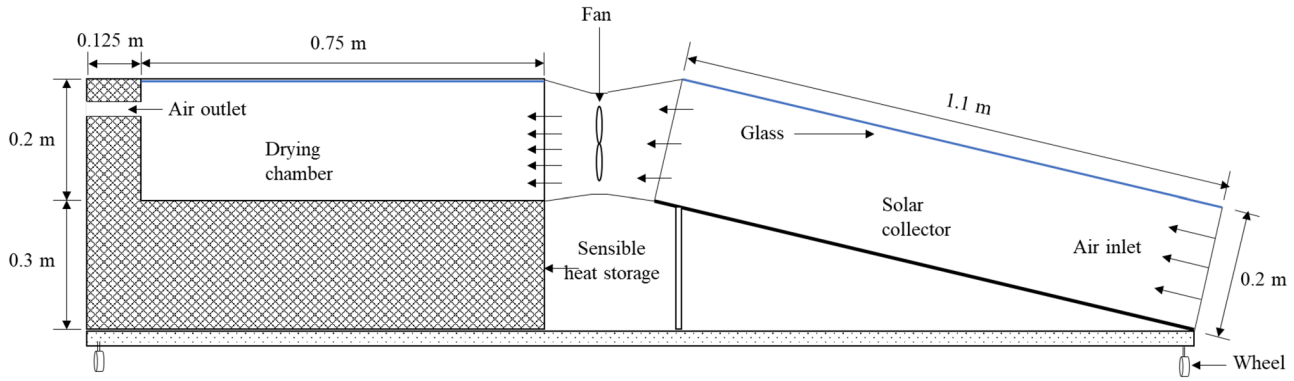


Fig. 30 Schematics of intermittent solar dryer [70]

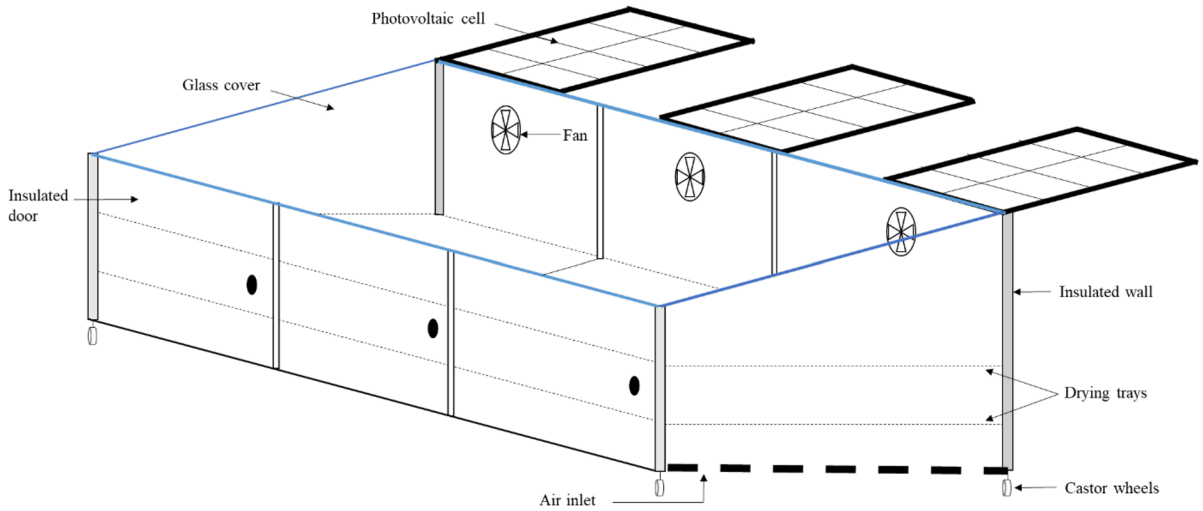


Fig. 31 Solar cabinet dryer [71]

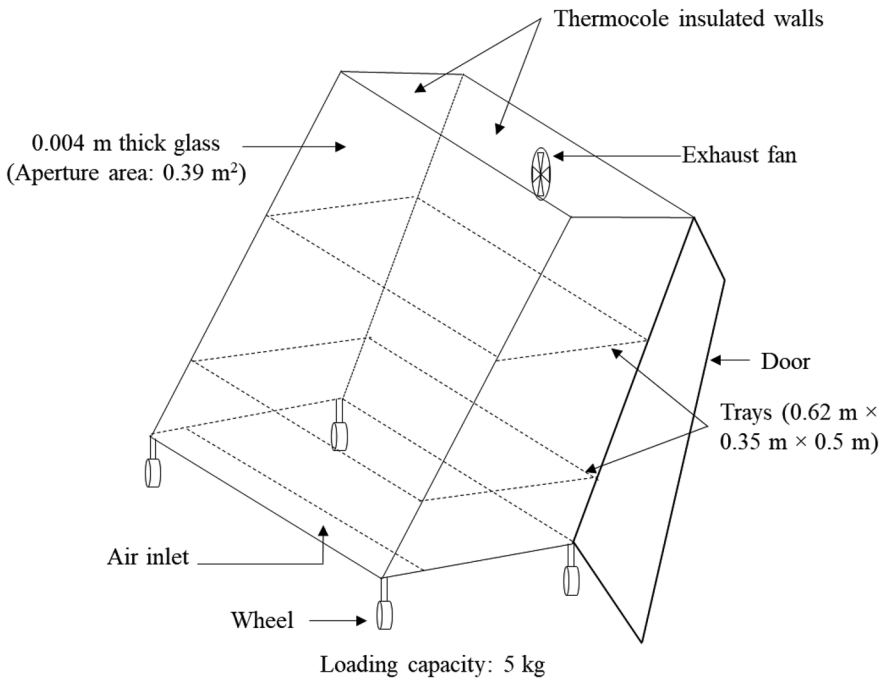


Fig. 32 Cabinet-type direct FCSD [72]

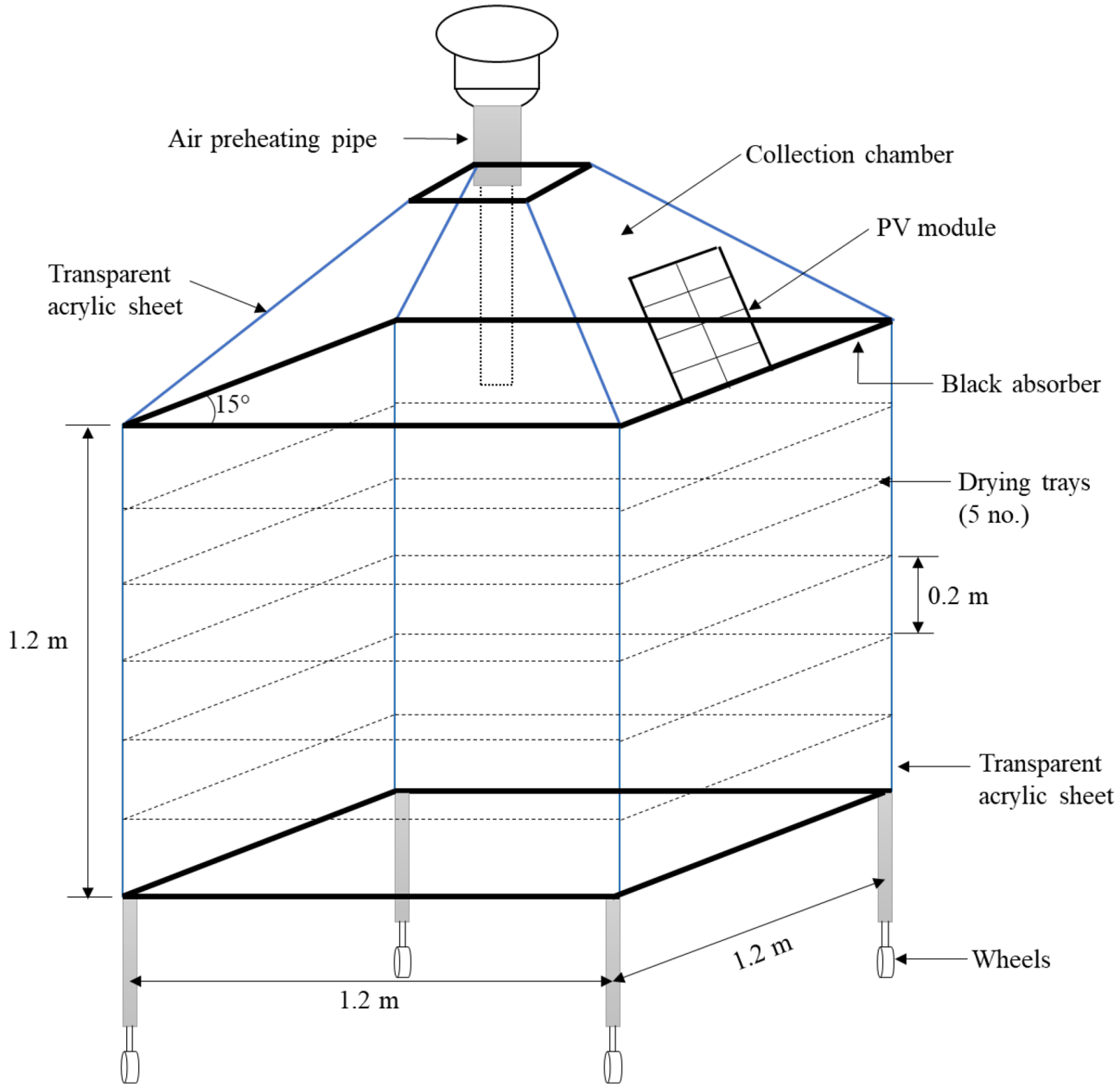


Fig. 33 Compact solar cabinet dryer [73]

further developments as there is a lack of such investigations on domestic solar dryers.

- Most of the research was confined to temperature variations and statistical presentations, but the dried product's quality, which is the most important factor in the whole process, has not even been discussed in maximum investigations.
- Environmental impact analysis is getting global recognition as one of the most important tools for presenting the sustainability potential of green technologies. There are only a few investigations where researchers have tried to analyze the environmental impact of a small-scale or domestic solar dryer.
- The quality of the dried product and the sustainability of the drying process with a simultaneous economic analysis should be presented in solar drying technologies.

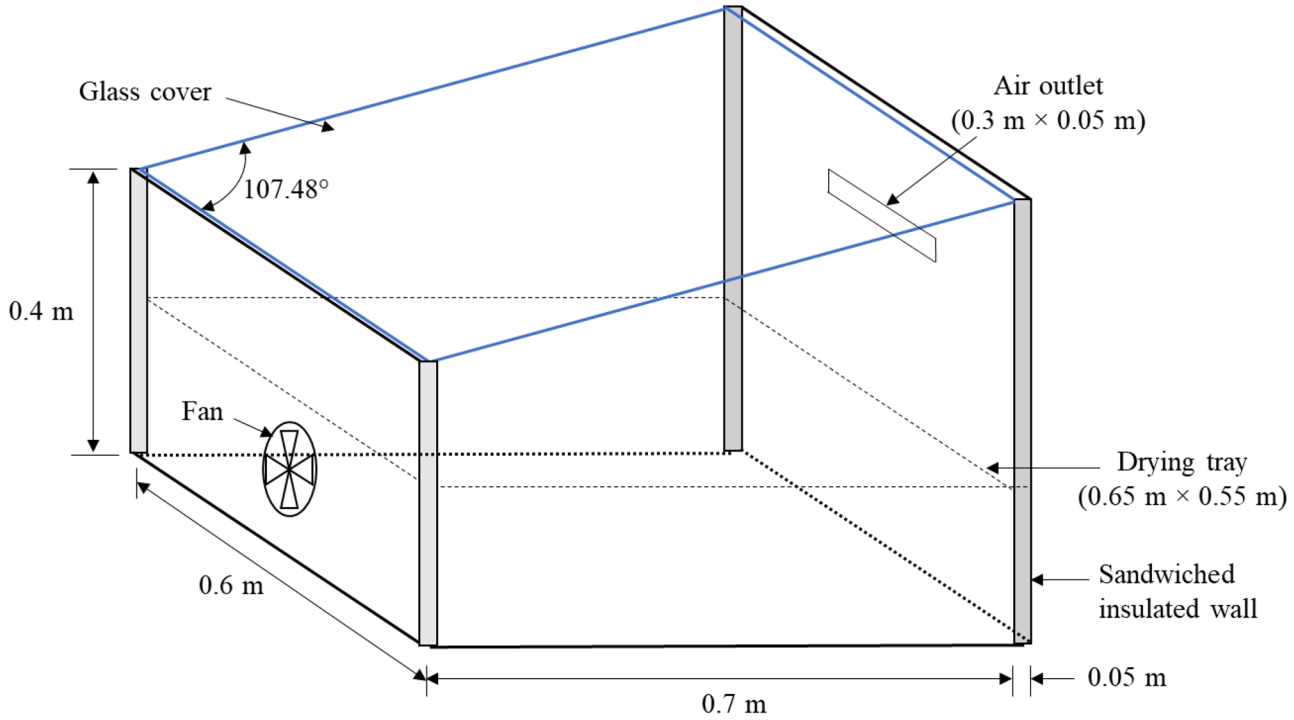


Fig. 34 Direct type solar fish dryer [74]

Table 6 Summary of domestic FCSD for miscellaneous commodities drying

Sr. No.	Researcher/s	Year	Country	Drying commodity	Description	Remarks
1	Fagunwa et al. [70]	2009	Nigeria	Cocoa beans	Mixed-mode solar dryer with an energy storage unit for the off-shine period was fabricated and tested	The highest moisture reduction of 53.4 to 3.56% (w.b.) in 72 h was found under free convection
2	Eswara and Ramakrishnarao [71]	2012	India	–	The application of solar technologies in food processing was presented	Solar technologies must be investigated in combinations rather than single unit testing
3	Seveda and Jhajharia [72]	2012	India	Large cardamom (Amomum subulatum)	Direct FCSD was fabricated and tested	Collector efficiency varied from 17.23 to 51.71%, with a reduction of 50% in the drying time
4	Jangsawang [73]	2017	Thailand	Raw pork	Cabinet type FCSD was fabricated and tested	A temperature of 45.75–48.63 °C was observed on different drying trays
5	Alonge and Obayopo [74]	2019	Nigeria	Fish	Direct FCSD was simulated using ANSYS FLUENT software and validated	Maximum collector efficiency and temperature elevations were 77.2% and 26.7 °C, respectively
6	Nimnuan and Nabnean [75]	2020	Thailand	Galangal (Alpinia galangal (Linn.) Swartz.)	Parabolic-shaped direct FCSD was tested	Drying time was reduced by 43% with an efficiency of 32%
7	Safri et al. [76]	2020	Malaysia	No-load	A potable greenhouse dryer was tested with four different solar air collector arrangements	The efficiency of flat plate and V-groove without insulator was 66.34%

Table 7 Economics of various domestic solar dryers

S. No.	Researcher/s	Year	Dryer	Commodity	Economic parameters	Remarks
1	Thanvi and Pande [29]	1987	NCSD	Chilies	The total fabrication cost of the dryer was estimated to be $22	Costs 50% less than a simple cabinet solar dryer
2	Singh et al. [32]	2006	NCSD	Chilies and fenugreek leaves	Capital and annualized cost of the NCSD was $35.55 and $2.12, respectively. Payback time was 0.57 and 1.36 years for chilies and fenugreek leaves	The estimated life was 20 years with 150 drying days per year. Drying cost of fenugreek leaves was 40% less than an electric dryer
3	Sreekumar et al. [65]	2008	FCSD	Bitter gourd	Total capital and annualized cost was $144.44 and $20.44, respectively. Payback time of the FCSD was 3.26 years	Drying cost was 42% less than an electric dryer. The estimated lifetime was 20 years with 250 working days per year
4	Alonge and Omoniwa [34]	2012	NCSD	Cassava chips	Production cost of the NCSD was $500	Total cost involved in the fabrication was presented
5	Tefera et al. [36]	2013	NCSD	Potato slices	Fabrication cost of the box and pyramid type NCSD was $39.92 and $10.94, respectively	A pyramid dryer was recommended due to its higher drying area, lesser weight and cost with the same drying time
6	Eke [39]	2014	NCSD	Tomato slices	The production costs for metal, cement, wood and mud dryer were $36.14, $25.22, $21.34 and $12.05, respectively	Production cost was highest for metal collector followed by cement, wood and mud collectors. However, metal dryer showed the highest efficiency
7	Mustapha et al. [54]	2014	NCSD	Fish	Fixed cost for different NCSD arrangements was varying from $11.25 to $21.25. Net income and cost benefit ratio were in the range of $28.12–$46.87 and 2.5:1–4.5:1. Payback period was estimated as 3 months	The lifetime for different solar dryers was in the range of 8–10 years
8	Modi et al. [67]	2017	FCSD	Tomato slices	Capital cost of the FCSD was $125.08 and the net profit was $4.7/ kg	Net profit was 6.4% higher as compared to OSD
9	Poonia et al. [26]	2018	NCSD	Ber (Ziziphus)	Total fabrication cost of the NCSD was $10.32	Low cost is more attractive for domestic and rural areas
10	Chaudhari et al. [43]	2018	NCSD	Ginger	Capital cost, net present worth, benefit cost ratio and payback time were $116.05, $266.92, 2.3 and 6 months, respectively	The daily benefit from drying was $0.89 and the lifetime of the dryer was 10 years with 300 operating days per year
11	Haque et al. [44]	2018	NCSD	Bitter gourd, okra, hirda and raw mango	Total fabrication and annualized cost were $77.36 and $12.12, respectively. Cumulative present worth was $1216.55 and the payback time was 0.56 years	Lifetime was considered as 10 years with 140 working days per year
12	Poonia et al. [61]	2018	FCSD	Ber (Zizyphus mauritiana)	Initial investment and net present worth of the FCSD was $180.53 and $477.33, respectively. Benefit–cost ratio, internal rate of return and payback period were 1.86, 54.5% and 2.26 years, respectively	A high internal rate of return and low payback time showed the economic suitability of the FCSD
13	Nabnean and Nimnuan [63]	2020	FCSD	Banana slices	Total installation cost and payback time of the FCSD were $390 and 1.1 years, respectively	The price of the solar-dried banana was 20% higher than OSD

Table 7 (continued)

S. No.	Researcher/s	Year	Dryer	Commodity	Economic parameters	Remarks
14	Nimnuan and Nabnean [75]	2020	FCSD	Galangal	Installation cost of the FCSD was $375 and payback time was 0.9 years	The return on solar-dried products is 30% higher than OSD
15	Sandali et al. [60]	2020	NCSD	–	Life cycle cost and benefit of the NCSD were $28,843.37 and $37,017.44, respectively. Payback time was estimated as 0.9 years, which is quite low	Working life was 15 years with 365 operating days
16	Moghimi et al. [69]	2021	FCSD	Tomato	The total cost of the FCSD was estimated as $309.52	The total development cost was very high which seems out of the reach of household users

Acknowledgements The authors are highly grateful to Guru Jambheshwar University of Science & Technology, Hisar, India and the Centre for Energy and Environment, Delhi Technological University, Delhi, India, for providing the necessary facilities to compile this work.

Funding No funding was received for conducting this study.

Declarations

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

References

1. Ekechukwu OV, Norton B (1999) Review of solar-energy drying systems II: an overview of solar drying technology. Energy Convers Manag 40:615–655. https://doi.org/10.1016/s0140-6701(99)97881-5
2. Lutz K, Mühlbauer W, Müller J, Reisinger G (1987) Development of a multi-purpose solar crop dryer for arid zones. Sol Wind Technol 4:417–424. https://doi.org/10.1016/0741-983X(87)90016-6
3. FAO (2019) The state of food and agriculture 2019. Moving forward on food loss and waste reduction. Rome
4. United Nations Environment Programme (2021) Food Waste Index Report 2021. Nairobi
5. NAAS (2019) Saving the harvest: reducing the food loss and waste. Policy Brief No. 5, National Academy of Agricultural Sciences, New Delhi
6. Stanley KE, Colo FC (1980) Solar food dryer. United States Patent. US4221059A
7. Kumar M, Sansaniwal SK, Khatak P (2016) Progress in solar dryers for drying various commodities. Renew Sustain Energy Rev 55:346–360. https://doi.org/10.1016/j.rser.2015.10.158
8. Shimpy, HM, Kumar M, Gupta M (2019) Recent developments and comprehensive review on greenhouse dryers. Trends and Advances in Mechanical Engineering (TAME 2019), Faridabad 23–31
9. El Hage H, Herez A, Ramadan M et al (2018) An investigation on solar drying: a review with economic and environmental assessment. Energy 157:815–829. https://doi.org/10.1016/j.energy.2018.05.197
10. Seerangurayar T, Al-Ismaili AM, Jeewantha LHJ, Al-Habsi NA (2019) Effect of solar drying methods on color kinetics and texture of dates. Food Bioprod Process 116:227–239. https://doi.org/10.1016/j.fbp.2019.03.012
11. Mulokozi G, Svanberg U (2003) Effect of traditional open sun-drying and solar cabinet drying on carotene content and vitamin A activity of green leafy vegetables. Plant Foods Hum Nutr 58:1–15. https://doi.org/10.1023/b:qual.0000041153.288
12. Navale SR, Harpale VM, Mohite KC (2015) Comparative study of open sun and cabinet solar drying for fenugreek leaves. Int J Renew Energy Technol Res 4:2325–3924
13. Belessiotis V, Delyannis E (2011) Solar drying. Sol Energy 85:1665–1691. https://doi.org/10.1016/j.solener.2009.10.001
14. Sukhatme SP, Nayak JK (2017) Solar Energy, 4th ed. McGraw-Hill Education, New Delhi
15. Chavan A, Vitankar V, Mujumdar A, Thorat B (2020) Natural convection and direct type (NCDT) solar dryers: a review. Dry Technol. https://doi.org/10.1080/07373937.2020.1753065
16. Prakash O, Kumar A, Sharaf-Eldeen YI (2016) Review on Indian solar drying status. Curr Sustain Energy Reports 3:113–120. https://doi.org/10.1007/s40518-016-0058-9
17. Yaciuk G (1981) Solar crop drying. Solar Energy Conversion II. Pergamon, Ontario, pp 377–339

18. Singh P, Gaur MK (2020) Review on development, recent advancement and applications of various types of solar dryers. Energy Sources A Recover Util Environ Eff 1–21. https://doi.org/10.1080/15567036.2020.1806951
19. Mohana Y, Mohanapriya R, Anukiruthika T et al (2020) Solar dryers for food applications: Concepts, designs, and recent advances. Sol Energy 208:321–344. https://doi.org/10.1016/j.solener.2020.07.098
20. Fournier M, Guinebault A (1995) The "shell" dryer-modelling and experimentation. Renew Energy 6:459–463
21. Hallak H, Hilal J, Hilal F, Rahhal R (1996) The staircase solar dryer: Design and characteristics. Renew Energy 7:177–183. https://doi.org/10.1016/0960-1481(95)00127-1
22. Wakjira M, Adugna D, Berecha G (2011) Determining slice thickness of banana (Musa spp.) for enclosed solar drying using solar cabinet dryer under ethiopian condition. Am J Food Technol 6:568–580. https://doi.org/10.3923/ajft.2011.568.580
23. Rawat BS, Rawat PN, Pant PC, Joshi GC (2014) Evaluation of energetics and CO2 emission mitigation potential of natural convection solar dryer for amla. AU J Technol 18:75–81
24. Terres H, Chavez S, Lopez R, et al (2015) Study of the lemon drying process using a solar dryer. In: ASME 2015 9th International Conference on Energy Sustainability. San Diego, California, pp 1–6
25. Daud LEI, Simate IN (2017) Drying kinetics of sliced pineapples in a solar conduction dryer. Energy Environ Res 7:14. https://doi.org/10.5539/eer.v7n2p14
26. Poonia S, Singh AK, Santra P, Mishra D (2018) Design, development and performance evolution of a low-cost solar dryer. In: Chandra L, Dixit A (eds) Concentrated Solar Thermal Energy Technologies, Springer Proceedings in Energy. Singapore, Springer Nature, pp 219–223
27. Islam MMI, Islam MMI, Tusar M, Limon AH (2019) Effect of cover design on moisture removal rate of a cabinet type solar dryer for food drying application. Energy Procedia 160:769–776. https://doi.org/10.1016/j.egypro.2019.02.181
28. Sharma VK, Sharma S, Ray RA, Garg HP (1986) Design and performance studies of a solar dryer suitable for rural applications. Energy Convers Manag 26:111–119. https://doi.org/10.1016/0196-8904(86)90040-3
29. Thanvi KP, Pande PC (1987) Development of a low-cost solar agricultural dryer for arid regions of India. Energy Agric 6:35–40. https://doi.org/10.1016/0167-5826(87)90020-9
30. Sharma S, Sharma VK, Jha R, Ray RA (1990) Evaluation of the performance of a cabinet type solar dryer. Energy Convers Manag 30:75–80. https://doi.org/10.1016/0196-8904(90)90016-R
31. Ezekoye BA, Enebe OM (2006) Development and performance evaluation of modified integrated passive solar grain dryer. Pacific J Sci Technol 7:185–190
32. Singh PP, Singh S, Dhaliwal SS (2006) Multi-shelf domestic solar dryer. Energy Convers Manag 47:1799–1815. https://doi.org/10.1016/j.enconman.2005.10.002
33. Rawat BS, Pant PC, Joshi GC (2009) Energetics study of a natural convection solar crop dryer. Int J Ambient Energy 30:193–198. https://doi.org/10.1080/01430750.2009.9675096
34. Alonge AF, Omoniwa AO (2012) Development and modification of a direct passive solar dryer. In: NABEC-CSBE/SCGAB 2012 Joint Meeting and Technical Conference Northeast Agricultural & Biological Engineering Conference Canadian Society for Bioengineering. Orillia, Ontario, pp 1–10
35. Alonge AF, Adeboye OA (2012) Drying rates of some fruits and vegetables with passive solar dryers. Int J Agric Biol Eng 5. https://doi.org/10.3965/j.ijabe.20120504.00
36. Tefera A, Endalew W, Fikiru B (2013) Evaluation and demonstration of direct solar potato dryer. Livest Res Rural Dev 25:1–7
37. Eke AB (2013) Development of small scale direct mode natural convection solar dryer for tomato, okra and carrot. Int J Eng Adv Technol 3:199–204
38. Raju RVS, Reddy RM, Reddy ES (2013) Design and fabrication of efficient solar dryer. J Eng Res Appl 3:1445–1458
39. Eke AB (2014) Investigation of low cost solar collector for drying vegetables in rural areas. Agric Eng Int CIGR J 16:118–125
40. Borah A, Hazarika K, Khayer SM (2015) Drying kinetics of whole and sliced turmeric rhizomes (Curcuma longa L.) in a solar conduction dryer. Inf Process Agric 2:85–92. https://doi.org/10.1016/j.inpa.2015.06.002
41. Subedi TR, Bhattarai RN (2017) Experimental performance analysis of solar conduction dryer (SCD) for ginger drying. In: IOE Graduate Conference. pp 597–601
42. Chaudhari RH, Bhavsar S (2017) Hybrid solar box type dryer cum cooker of chilly drying for domestic usage. Int J Sci Res 6:1614–1618
43. Chaudhari RH, Gora A, Modi VM, Chaudhari H (2018) Economic analysis of hybrid solar dryer for ginger drying. Int J Curr Microbiol Appl Sci 7:2725–2731. https://doi.org/10.20546/ijcmas.2018.711.312
44. Haque T, Tiwari M, Bose M, Kedare SB (2018) Drying kinetics, quality and economic analysis of a domestic solar dryer for agricultural products. Ina Lett 4:147–160. https://doi.org/10.1007/s41403-018-0052-1
45. Ugwuoke IC, Ikechukwu IB, Ifianyi OE (2019) Design and development of a mixed-mode domestic solar dryer. Int J Eng Manuf 9:55–65. https://doi.org/10.5815/ijem.2019.03.05
46. Chavan A, Vitankar V, Thorat B (2020) CFD modeling and experimental study of solar conduction dryer. Dry Technol 39:1087–1100. https://doi.org/10.1080/07373937.2020.1846051
47. Ampratwum DB, Dorvlo ASSS (1998) Evaluation of a solar cabinet dryer as an air-heating system. Appl Energy 59:63–71. https://doi.org/10.1016/S0306-2619(97)00043-3
48. Bolaji B (2005) Development and performance evaluation of a box-type absorber solar air collector for crop drying. J Food Technol 3:595–600
49. Mwithiga G, Kigo SN (2006) Performance of a solar dryer with limited sun tracking capability. J Food Eng 74:247–252. https://doi.org/10.1016/j.jfoodeng.2005.03.018
50. Kumar N, Agravat S, Chavda T, Mistry HN (2008) Design and development of efficient multipurpose domestic solar cookers/dryers. Renew Energy 33:2207–2211. https://doi.org/10.1016/j.renene.2008.01.010
51. Saleh A, Badran I (2009) Modeling and experimental studies on a domestic solar dryer. Renew Energy 34:2239–2245. https://doi.org/10.1016/j.renene.2009.03.001
52. Abdullahi Y, Momoh M, Garba MM, Musa M (2013) Design and construction of an adjustable and collapsible natural convection solar food dryer. Int J Comput Eng Res 3:1–8
53. Ozuomba JO, Okonkwo NA, Uzor BC, Uba JI (2013) Fabrication and characterization of a direct absorption solar dryer. Adv Appl Sci Res 4:186–194
54. Mustapha MK, Salako AF, Ademola SK, Adefila IA (2014) Qualitative performance and economic analysis of low cost solar fish driers in Sub-Saharan Africa. J Fish 2:64
55. Akpojaro J, Oyeyemi M (2015) Performance evaluation of a prototype solar dryer against the conventional sun-drying system in Nigeria. Br J Appl Sci Technol 9:411–418. https://doi.org/10.9734/bjast/2015/16308
56. Chavan A, Sikarwar A, Tidke V, Thorat B (2018) Augmenting natural convection and conduction based solar dryer. 21st International Drying Symposium. València, Spain, pp 1357–1364
57. Sandali M, Boubekri A, Mennouche D (2018) Thermal behavior modeling of a cabinet direct solar dryer as influenced by sensible

heat storage in a fractured porous medium. AIP Conf Proc 1968. https://doi.org/10.1063/1.5039173
58. Mehata P, Jhala R, Harichandan A (2018) Design and mathematical modelling of mixed mode solar dryer applicable for small scale application. Int J Eng Res Mech Civ Eng 3:173–177
59. Jain A, Sharma M, Kumar A et al (2019) Computational fluid dynamics simulation and energy analysis of domestic direct-type multi-shelf solar dryer. J Therm Anal Calorim 136:173–184. https://doi.org/10.1007/s10973-018-7973-5
60. Sandali M, Boubekri A, Mennouche D (2020) Thermal and economical study of a direct solar dryer with integration of different techniques of heat supply. In: Belasri A, Beldjilali S (eds) ICREEC 2019. Springer Proceedings in Energy. Springer, Singapore. https://doi.org/10.1007/978-981-15-5444-5_73
61. Poonia S, Singh AK, Jain D (2018) Design development and performance evaluation of photovoltaic/thermal (PV/T) hybrid solar dryer for drying of ber (Zizyphus mauritiana) fruit. Cogent Eng 5:1–18. https://doi.org/10.1080/23311916.2018.1507084
62. Poonia S, Singh AK, Jain D (2018) Mathematical modelling and techno-economic evaluation of hybrid photovoltaic-thermal forced convection solar drying of Indian jujube (Zizyphus mauritiana). J Agric Eng 55:74–88
63. Nabnean S, Nimnuan P (2020) Experimental performance of direct forced convection household solar dryer for drying banana. Case Stud Therm Eng 22:100787. https://doi.org/10.1016/j.csite.2020.100787
64. Tiris C, Tiris M, Dincer I (1996) Experiments on a new small-scale solar dryer. Appl Therm Eng 16:183–187
65. Sreekumar A, Manikantan PE, Vijayakumar KP (2008) Performance of indirect solar cabinet dryer. Energy Convers Manag 49:1388–1395. https://doi.org/10.1016/j.enconman.2008.01.005
66. Gaikwad SS, Shinde AB, Mote AA, Kachare PS (2016) Design and construction of briefcase type portable solar dryer. In: Pawar PM et al (eds) Techno-Societal 2016. Springer, Cham, pp 771–778
67. Modi VM, Desai NN, Gora A (2017) Design and development of low cost solar dryer. AGRES - An Int e J 6:329–336
68. Mohsen HA, Abd El-Rahmam AA, Hassan HE (2019) Drying of tomato fruits using solar energy. Agric Eng Int CIGR J 21:204–215
69. Moghimi P, Rahimzadeh H, Ahmadpour A (2021) Experimental and numerical optimal design of a household solar fruit and vegetable dryer. Sol Energy 214:575–587. https://doi.org/10.1016/j.solener.2020.12.023
70. Fagunwa AO, Koya OA, Faborode MO (2009) Development of an intermittent solar dryer for cocoa beans. Agric Eng Int CIGR J 11:1–14
71. Eswara AR, Ramakrishnarao M (2013) Solar energy in food processing - a critical appraisal. J Food Sci Technol 50:209–227. https://doi.org/10.1007/s13197-012-0739-3
72. Seveda MS, Jhajharia D (2012) Design and performance evaluation of solar dryer for drying of large cardamom (Amomum subulatum). J Renew Sustain Energy 4. https://doi.org/10.1063/1.4769199
73. Jangsawang W (2017) Meat products drying with a compact solar cabinet dryer. Energy Procedia 138:1048–1054. https://doi.org/10.1016/j.egypro.2017.10.103
74. Alonge OI, Obayopo SO (2019) Computational fluid dynamics and experimental analysis of direct solar dryer for fish. Agric Eng Int CIGR J 21:108–117
75. Nimnuan P, Nabnean S (2020) Solar drying of galangal slices (alpinia galangal (linn.) swartz.) using household solar dryer. Suranaree J Sci Technol 27:1–8
76. Safri NAM, Zainuddin Z, Mohd Azmi MS et al (2020) Temperature performance of a portable solar greenhouse dryer with various collector design. Sains Malaysiana 49:2539–2545. https://doi.org/10.17576/jsm-2020-4910-19

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.

Ain Shams Engineering Journal xxx (xxxx) xxx

Contents lists available at ScienceDirect

Ain Shams Engineering Journal

journal homepage: www.sciencedirect.com

Dielectric frequency response at different voltage levels of mineral oil-based silica nanofluids

H.C. Verma [a,*], A. Baral [b]

[a] Department of Electrical Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, Jharkhand 826004, India
[b] Department of Electrical Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, Jharkhand 826004, India

ARTICLE INFO

Article history:
Received 6 July 2022
Revised 31 August 2022
Accepted 21 September 2022
Available online xxxx

Keywords:
Dielectric frequency response
Silica Nanofluid
Voltage level
Temperature
Oil-paper composite insulation

ABSTRACT

In this paper, the Frequency domain dielectric response of oil-paper insulation, employing SiO_2-based nanofluids is reported. It is a well-known fact that the dissipation factor ($tan\delta$) profile is affected by the magnitude of excitation voltage & measurement temperature. Analysis of data obtained from laboratory samples suggests that the influence of excitation voltage is non-linear in nature. It is further observed that the degree of non-linearity is different for new and aged mineral oil. The frequency domain dielectric response for new mineral oil-based silica nanofluid has been found non-linearity nature at 50 Hz for sinusoidal excitation level 26 V-130 V for temperature range 50°C-70°C. Similarly, in the case of aged mineral oil, a non-linearity nature has been found at 1 Hz. The related analysis presented in this paper suggests that it is possible to express the non-linearity introduced in the value of dissipation factor ($tan\delta$) at power frequency by a mathematical expression.

© 2022 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Ain Shams University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

With the ever-increasing demand for power transmission and distribution, improving the dielectric properties of oil-paper insulation has become a focal point for researchers and power engineers. The principal advantage of oil-paper insulation is that cellulosic materials (i.e., kraft paper, pressboard, etc.) are cheap and abundant. However, to increase the power handling capacity, it is required that the oil-paper insulation should be less bulky and have higher dielectric strength. Transformer oil-based nanofluids show great promise in fulfilling insulation requirements [2–6]. Choi [4] first introduced the term nanofluid at Argonne National Laboratory of the USA in 1995. Nanofluid is a new class of fluid consisting of Nano-sized particles (1–100 nm) suspended with base fluid. Nanoparticles are mainly metal, metal oxide, which increases convection and conduction coefficients [8]. So, the insulation and coolant properties of fluid increase. Nanofluids have shown significant improvements in dielectric and thermal properties. S. Maneerot et al. [19] show improvement in electrical properties and also discuss dielectric loss factor ($tan\delta$) at 50 Hz for TiO_2-based nanofluid. The literature [20] reviews synthesis, experimental investigation and dielectric & physio-mechanical characterization of nanotechnology-based insulating liquids.

Over the past few years, researchers have opted for conductive (Fe_3O_4, Fe_2O_3, ZnO, SiC), semiconductive (TiO_2, CuO, Cu_2O) and insulating (Al_2O_3, SiO_2) nanoparticle-based mineral oil, and Natural ester for improving performance enhancement of power transformer [7,8]. Y. Z. Lv et al. [1] reported that conductive nanoparticle (Fe_3O_4) based fluid has greater dielectric strength, cooling property, and heat transfer property compared to base fluid mineral oil. Several publications [1,3,9,10] show that maximum enhancement of AC breakdown voltage has been observed when Fe_3O_4, TiO_2 nanoparticle-based nanofluid is used. H. Jin et al. [9] four different % volume concentrations of silica nanofluid were prepared, and it was found that the maximum 17 % enhancement of AC breakdown voltage when silica nanoparticle is mixed with base fluid. Metallic oxide Nanoparticles are mainly used due to their low cost, easy fabrication, and stable nanofluid [11].

Available literature [1,2,17] has shown that SiO_2 nanoparticle-based mineral oil can improve 17 % AC breakdown voltage when 0.0074 % volume concentration of SiO_2 nanoparticle is used. Literature [21,22,23] & [24] have shown that SiO_2 nanoparticle-based

* Corresponding author.
E-mail address: hcverma1989@gmail.com (H.C. Verma).

Peer review under responsibility of Ain Shams University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.asej.2022.101994
2090-4479/© 2022 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Ain Shams University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article as: H.C. Verma and A. Baral, Dielectric frequency response at different voltage levels of mineral oil-based silica nanofluids, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2022.101994

mineral oil can improve electrical, physical and chemical properties. Silicon dioxide (SiO_2) is used because of its electron quenching, free radical scavenging mechanisms and better performance in terms of its hydrophilicity, which sequestrates water molecules. In this work, the effect of voltage and temperature on the dielectric frequency response of silica-based nanofluid-impregnated oil-paper insulation has been studied. For this purpose, FDS studies were performed at five different sinusoidal excitation voltages (20–130 V) at 50°C, 55°C, 60°C, and 70°C temperatures on silica-nanofluid and mineral oil-based oil-paper insulation. It was observed that, unlike mineral oil-based insulation, *tanδ* characteristics of silica-nanofluid insulation are highly dependent on temperature and voltage at 50 *Hz*. This non-linear nature of silica nanofluid-based insulation was studied in detail because this may limit the application of silica-based nanofluids in practical transformers, as tanδ at 50 Hz is a reliable insulation performance parameter used by power utilities and a methodology for compensating the effect of voltage and temperature is described.

2. Experimental setup

2.1. Preparation of nanofluids

There are two methods to prepare nanofluid, namely-one-step and two-step methods [8]. The one-step method is mainly a chemical process in which the preparation of nanoparticles and dispersion in base fluid occur simultaneously. So incomplete reaction or stabilization with base fluid restricts the applicability of such nanofluid. In the two-step method, Nanoparticles are prepared using a physical or chemical process [1]. Magnetic stirrer, Ultrasonic Bath is used to disperse nanoparticle with base fluid. The nanoparticles used in the present work were obtained from Nano Labs Jamshedpur, Jharkhand, India. As per the report obtained from Nano-labs, the SiO_2 used for the present work is 99.5 % pure with a particle size range of 20–30 nm having a density of 2.5 g/cm^{3}and the average particle size is 20 nm. The following steps were used to prepare Nanofluids, which are shown in Fig. 1:

1 The SiO_2 Nanoparticles are weighted and mixed with 1000 ml of mineral oil.

2. First, Magnetic stirring (shown in Fig. 2(a)) is used to disperse the nanoparticle with mineral oil for 15 min. After magnetic stirring, the mixture was ultrasonically treated for 30 Minutes at 30°C for complete dispersion of SiO_2 nanoparticles with mineral oil [18].

In the present work, 0.0074 % volume concentration silica Nanofluid was tested throughout the experiments, as [1] and [12] suggested, that this particular prepared for the volume concentration of SiO_2 with mineral oil base fluid leads to a 17 % improvement of AC Breakdown voltage.

FESEM analysis of nanoparticle SiO_2 was done to collect information about nanoparticle uniformity. It is found that its characteristics are wet, uncoated, porous and no surfactant added. The analysis also shows particle sphericity, the solution stability, and its homogeneity. A pictorial view of SiO_2 nanoparticles is shown in Fig. 3.

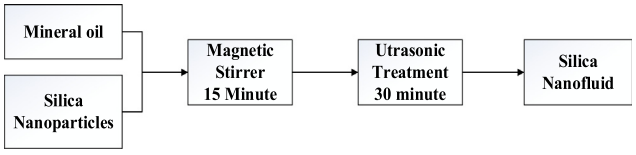


Fig. 1. The procedure of preparation of silica-based Nanofluid.

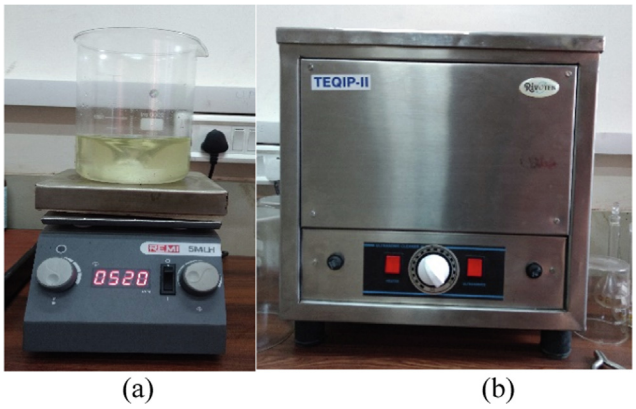


Fig. 2. Prepared test samples in the laboratory. (a) Magnetic stirrer and (b) Ultrasonic Bath.

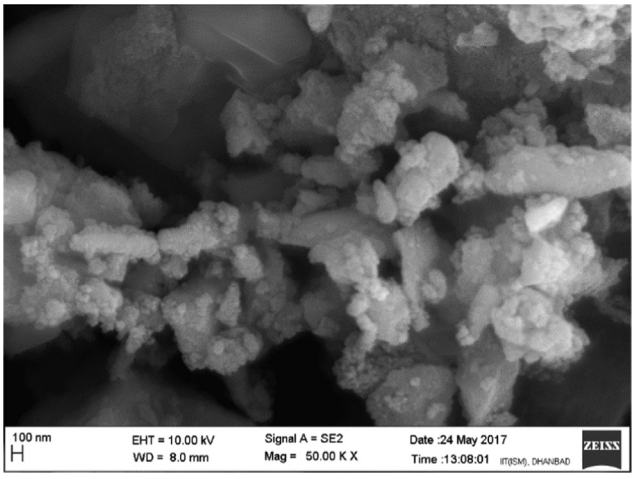


Fig. 3. FESEM micrograph for 50,000 times zooms in the scale of 100 nm.

2.2. Laboratory sample

For dielectric frequency response, a test sample has been prepared in a laboratory that emulates the oil-paper insulation of a real power transformer. A test sample is constructed using four mils thick Kraft paper, pressboard, and copper sheet. Its view is shown in Fig. 4(a). The detailed procedure of the construction of a test sample is discussed in the papers [10,15]. Fig. 4(b) shows the constructed sample. For removing the moisture from the constructed test sample, the sample is heated continuously at 90 °C for 96 h shown in Fig. 5(a) and 5(b). Another method is used for removing moisture from the constructed sample. The Constructed test sample was put in a temperature-controlled oven set at 80 °C to ensure that it was dry. After being heated for 72 h, it was observed that the test sample was more than 97 % dry. The

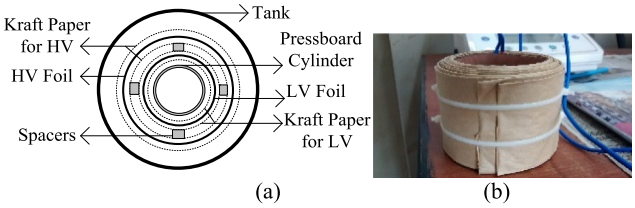


Fig. 4. Prepared test samples in the laboratory. (a) Sectional drawing and (b) actual photograph.

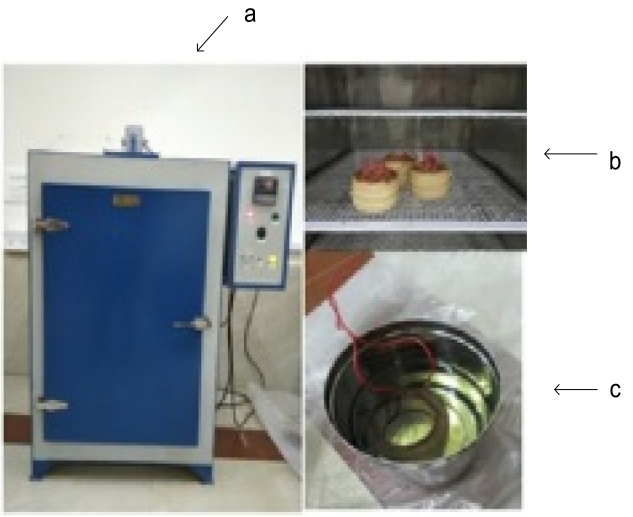


Fig. 5. a) Oven, b) Dry Sample, c) Impregnated with transformer oil.

thermo-gravimetric analyser (Citizen make; model MB50) was used to test the dryness of the test sample. Further, it was observed that after 72 h, the dry sample weight becomes approximately constant (tested using a precision balance; Citizen model CY1000 C). The samples were heated for a further 24 h at 80 °C to ensure their complete dryness. It is reported in [25] that the sample was heated for 96 h at 80 °C to ensure dryness of the sample. It is also mentioned in [25] that the dryness of constructed sample (heating temperature and duration) depends on the dimension and volume of paper present in the said sample.

After sufficient heating, a moisture-free test sample was obtained, and its weight (W_a) was measured with a precision weighing machine. For sufficient absorption of moisture by the constructed test samples, it was exposed to open air. After absorbing a pre-determined amount of moisture, its weight was found to be (W_b). The moisture content of the test samples was evaluated using equation (1) and was found to be 3 % [16].

$$Themoisturecontentofthesample = \frac{W_b - W_a}{W_a} \times 100\% \quad (1)$$

After absorbing a pre-determined amount of moisture content, the test sample was immersed in a steel container containing 1000 ml new mineral oil shown in Fig. 5(c) and also prepared SiO_2-based nanofluid.

2.3. Background of fds theory

In a practical transformer, FDS is carried out by connecting all HV bushings to each other, as are all LV bushings. Then excitation voltage of different frequencies is applied across HV and LV bushings. This initiates the polarization process inside insulation material resulting in the flow of current through it. The phase difference between the excitation voltage and the response current is used for the computation of various frequency-dependent performance parameters. The typical arrangement followed for the measurement of FDS data is illustrated in Fig. 6.

In the Laplace domain, the relation between the response current $I(\omega)$ and the excitation voltage $u(\omega)$ having a frequency f can almost be described by equation (2).

$$I(\omega) = j\omega c(\omega)u(\omega) \quad (2)$$

$$c(\omega) = c'(\omega) - jc''(\omega) \quad (3)$$

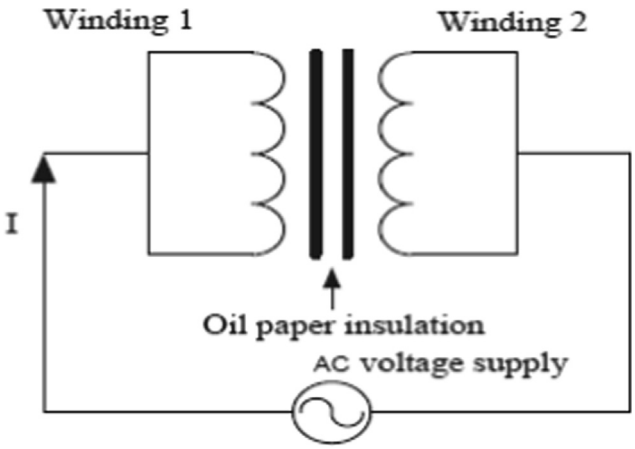


Fig. 6. FDS setup for Power Transformer.

the complex capacitance of the insulation system under test, which is correlated to the dielectric permittivity of insulating material. It can be observed that oil-paper insulation being a dielectric material, has both energy storing and energy dissipating capability. Hence, the dissipation factor $tan\delta$ can be written as the ratio of the imaginary part of complex capacitance to the real part of the complex capacitance (4).

$$\tan\delta = \frac{c''(\omega)}{c'(\omega)} \quad (4)$$

2.4. Fds experimental set-up

The schematic Diagram of the FDS measurement is shown in Fig. 7. The value of $tan\delta$ can be determined if information regarding the phase difference between excitation voltage and a current flowing through the circuit is known. Both excitation voltage and a current flowing through circuit data are saved. Using the synchronous detection in MATLAB programming, $tan\delta$ was calculated from the frequency range of 0.04642 *Hz* to 1*KHz*. In this setup, high noise current flows through the circuit when it goes lower than 0.0462 *Hz*. So in this paper, $tan\delta$ has been calculated up to a lower frequency range of 0.04642 *Hz*.$tan\delta$ at a different frequency range of laboratory test sample was measured at different sinusoidal excitation levels and temperature using a portable setup fabricated



Fig. 7. FDS measurement unit.

and designed at High voltage laboratory in IIT(ISM) Dhanbad shown in Fig. 7.

3. Fds measurement for laboratory sample

The frequency-domain dielectric response of a laboratory test sample containing 3 % moisture content for pure mineral oil has been measured at sinusoidal excitation levels between 26 V and 130 V for temperature ranges 50°C, 55°C, 60°C, 70°C shown in Fig. 8. For pure mineral oil, frequency domain dielectric response is measured at different temperatures with different voltage overlaps with each other. It shows that the dielectric frequency response characteristics of a mineral oil-impregnated sample do not depend much on temperature, particularly in the 1–1000 *Hz* frequency range, because of the lower value of activation energy.

3.1. Dielectric frequency response for new nanofluid-based oil-paper insulation at different voltage and different temperature

The frequency-domain dielectric response of laboratory test samples containing 3 % moisture content for new mineral oil-based silica nanofluid has been measured at sinusoidal excitation levels 26 V, 52 V, 78 V, 104 V, 130 V for temperature ranges 50°C, 55°C, 60°C, 70°C shown in Fig. 9. After the addition of silica with base fluid, frequency domain dielectric response has been non-linearity nature at a higher frequency range 5 *Hz*-1000 *Hz* shown in Fig. 9. In a new silica-based nanofluid, the variation of minimum *tanδ* lies across the higher frequency range 5 *Hz*-1000 *Hz,* so it improves dielectric insulation properties.

Individual dielectric frequency response for different voltage levels and the different temperatures is shown in Figure 10a, 10b, 10c and 10d. The value of minimum *tanδ* increases with increasing temperature. It is reported in [14] that the value of *tanδ* increases in the entire frequency range after the addition of insulating nanoparticles with mineral oil shown in Fig. 9. Figure 10a through Figure 10d reports the experimental results that are obtained for the nanofluid. In Figure 10a, the variation in dissipation factor (w.r.t that obtained for 26 V) at 50C is observed to be dominant between 1 Hz and 100 Hz. It is true that the variation observed in Fig. 10(a) is lesser than that observed in Fig. 10(b) though Figure 10d. With increase in temperature the variation in tanδ is found to increase and the frequency range at which this variation is observed is also found to shift right.

Dissipation factor denotes the loss that occurs within the insulation when the dipoles orient themselves in the direction of applied field. In the case of an ac field, the dipoles oscillate about their mean position which contributes to the tanδ value. With increase in temperature, the thermal energy content of the dipoles increases which in turn increase the inter-dipole interaction and the dipole-nanoparticles interaction during their oscillation. Furthermore, the thermal characteristics of the nanoparticle also play a role.

As per the supplier, the nanoparticles used for preparation of nanofluid have average particle size (APS) between 20 and

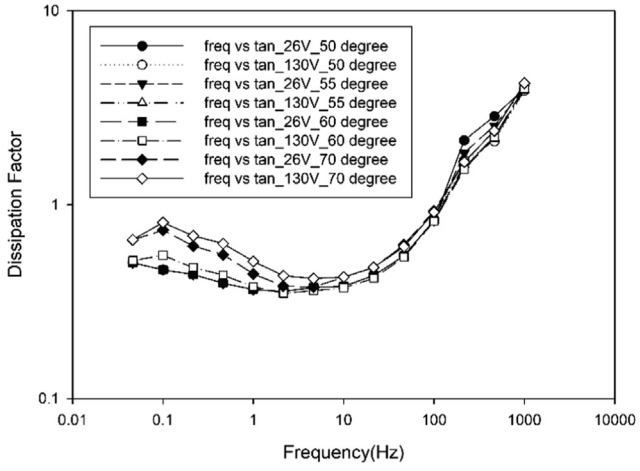


Fig. 8. Dielectric frequency response with different excitation for pure mineral oil at 50, 55, 60, & 70 °C.

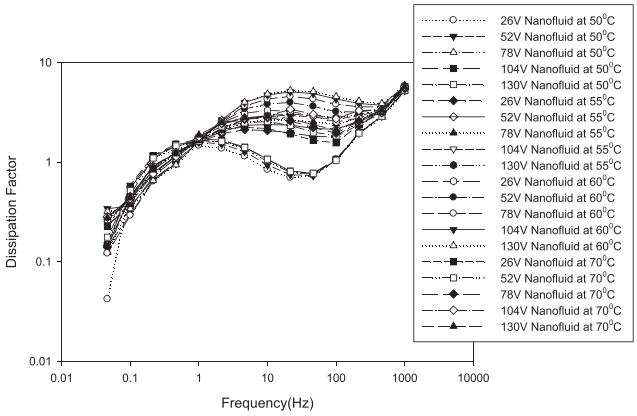


Fig. 9. Dielectric frequency response with different excitation nanofluid at 50 °C,55 °C,60 °C, and 70 °C.

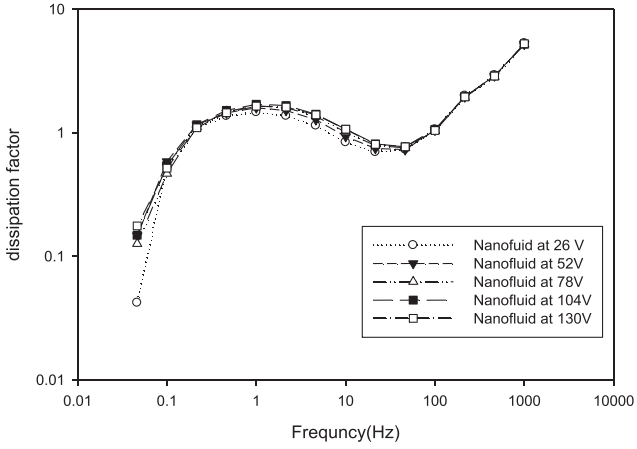


Fig. 10a. Dielectric frequency response with different excitation nanofluid at 50 °C.

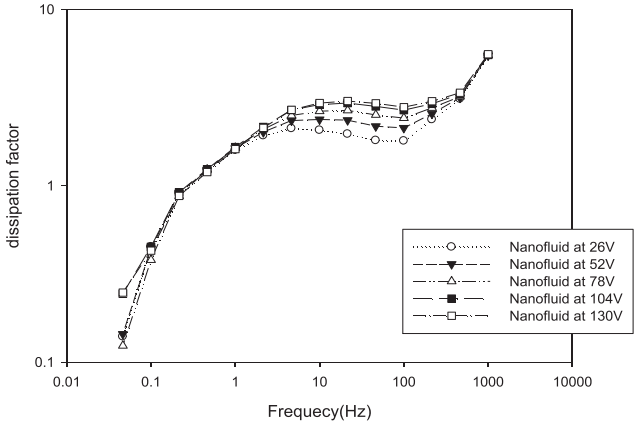


Fig. 10b. Dielectric frequency response with different excitation nanofluid at 55 °C.

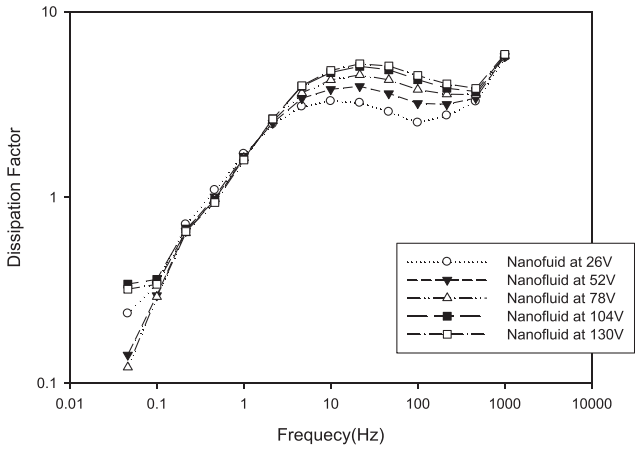


Fig. 10c. Dielectric frequency response with different excitation nanofluid at 60 °C.

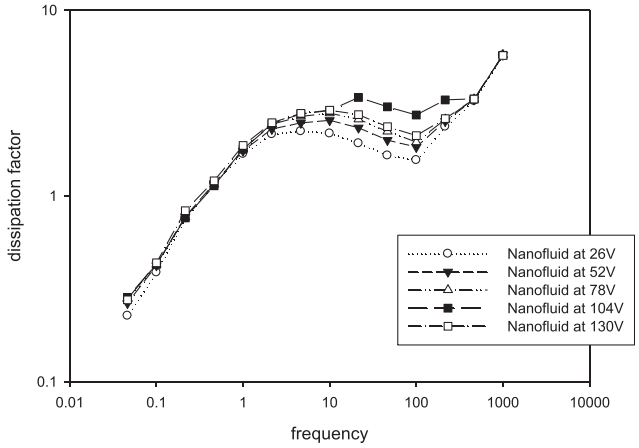


Fig. 10d. Dielectric frequency response with different excitation nanofluid at 70 °C.

40 nm. The variation observed in tanδ profile at a given temperature will be affected by several factors like type of nanoparticle used, volume concentration, APS, thermal characteristics of the nano particles used. At present detailed study regarding the influence of these factors on tanδ profile is being studied. The result obtained will be reported in future communications.

Figure 10a reports the variation of tanδ profile for different voltage magnitude that was observed for the tested nanofluid. Dissipation factor denotes the loss that occurs within the insulation when the dipoles orient themselves in the direction of applied field. In the case of an ac field, the dipoles oscillate about their mean position which contributes to the tanδ value. The dissipation factor profile at a given temperature depends on the thermal energy content of the dipoles thermal characteristics of the nanoparticle used.

At low frequencies, the dipoles residing in paper becomes active owing to limited mobility. As frequency increases the dipoles present in paper becomes less responsive. In other words, with increase in frequency the contribution of slow-moving dipoles to tanδ decreases. A reverse situation happens for fast moving dipoles that reside in bulk oil volume. As frequencies increase the contribution of contribution of fast-moving dipoles to tanδ increases. As both fast- and slow-moving dipoles co-exist in the case of oil-paper insulation, the overall profile of the tanδ depends on the interaction of these two dipole groups with the applied ac field. According to available literature, the dipoles residing at the oil-paper interface becomes dominant at 100 Hz [13]. This implies both dipoles residing at paper and oil will affect the value of tanδ at 100 Hz.

It is worth mentioning here that the authors of the paper have analyzed insulation response of several in-service units in the past. Such dip is generally not observed at such high frequency in the case of oil-paper insulation employing pure mineral oil. Hence, it can be opined that the dip in dissipation factor observed at 100 Hz along with the overall variation observed in tanδ profile at a given temperature is the result of interaction between several factors which include type of nanoparticle used, volume concentration, APS, thermal characteristics of the nano particles used. At present detailed study regarding the influence of these factors on tanδ profile is being studied. The result obtained will be reported in future communications.

3.2. Dielectric frequency response for aged nanofluid-based oil-paper insulation at different voltage and different temperature

The dielectric frequency response of the test sample containing 3 % moisture content for aged mineral oil-based silica nanofluid has been measured at 26 V, 52 V, 78 V, 104 V, 130 V excitation levels for 50°C, 55°C, 60°C, 70°C temperature shown in Fig. 11. After mixing silica with aged nanofluid, dielectric frequency response has been overlapping between 5 *Hz* and 464.15 *Hz* shown in Fig. 11. It is reported in [13] that if minimum *tanδ* shifts higher frequency range, then dielectric insulation of oil-paper insulation improves. In aged silica-based nanofluid, the variation of minimum

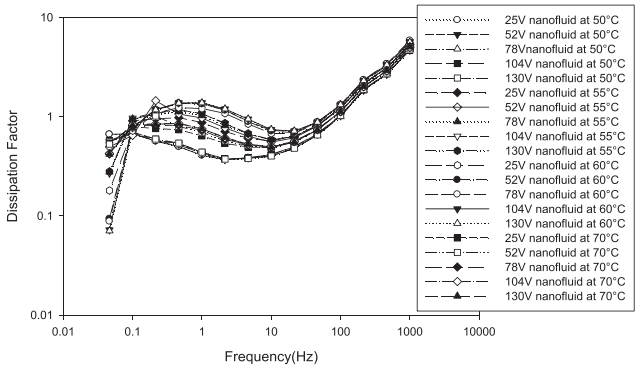


Fig. 11. Dielectric frequency response with different excitation nanofluid at 50 °C,55 °C,60 °C, and 70 °C.

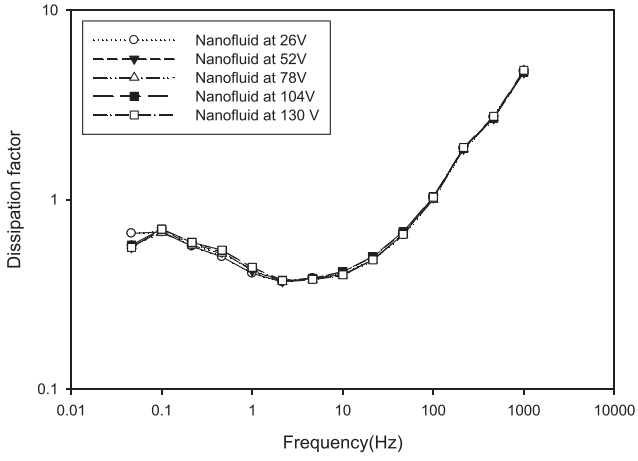


Fig. 12a. Dielectric frequency response with different excitation nanofluid at 50 °C.

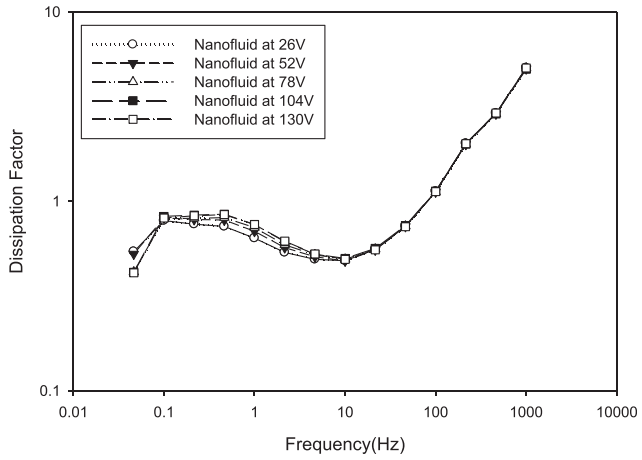


Fig. 12b. Dielectric frequency response with different excitation nanofluid at 55 °C.

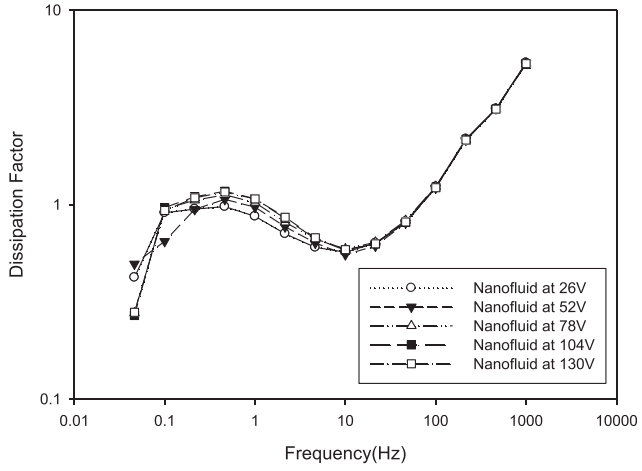


Fig. 12c. Dielectric frequency response with different excitation nanofluid at 60 °C.

tanδ lies across the lower frequency range 5 *Hz*-464.15 *Hz,* so it deteriorates dielectric insulation properties. That is why again, a new mineral oil-based nanofluid has been prepared.

Individual dielectric frequency responses for different voltage levels and different temperatures are shown in Figure 12a, 12b, 12c and 12d. It is reported in [13] that if minimum *tanδ* shifts higher frequency range, then dielectric insulation of oil-paper insulation improves. The value of minimum *tanδ* increases with increasing temperature.

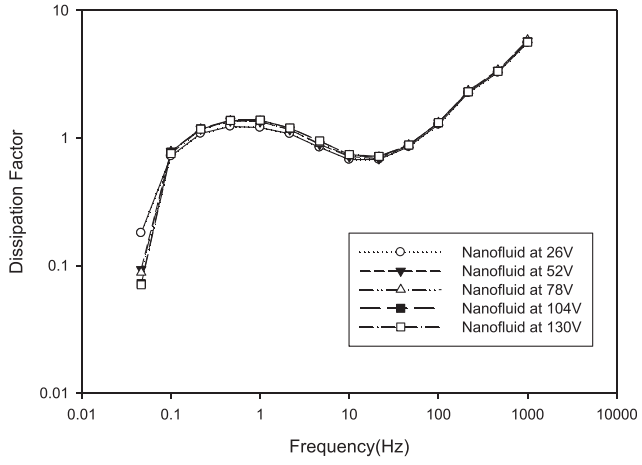


Fig. 12d. Dielectric frequency response with different excitation nanofluid at 70 °C.

4. Calculation of dissipation factor

4.1. For new mineral oil at constant frequency (50hz)

The deviation of *tanδ* at different temperatures is calculated from Figure 13a, 13b and 13c. The value of *Δtanδ (V)* at 50°C is fitted with the sigmoidal function shown in Figure 13a. Similarly, the value of *Δtanδ (V)* at 55°C, 60°C, and 70°C is also fitted with the sigmoidal function shown in Figure 13b, 13c and 13d. After the fitting process for different temperatures, the value of *Δtanδ (V, T)* at power frequency (50 Hz) is calculated using a combined sigmoidal curve fitting for all temperatures, which forms equation (5). Equation (5) is a sigmoidal function in which V is a different excitation level provided by the function generator.

If the *tanδ* values recorded at the same temperature under different excitation voltages are compared, a non-linear relationship is observed. This is evidenced in figure Figure 10a, 10b, 10c and 10d. This relationship can be given by the following equation,

$$\Delta \tan\delta(V,T) = \left(\frac{a(T)}{1+\exp\left(X_0(T) - V/b(T)\right)}\right) \quad (5)$$

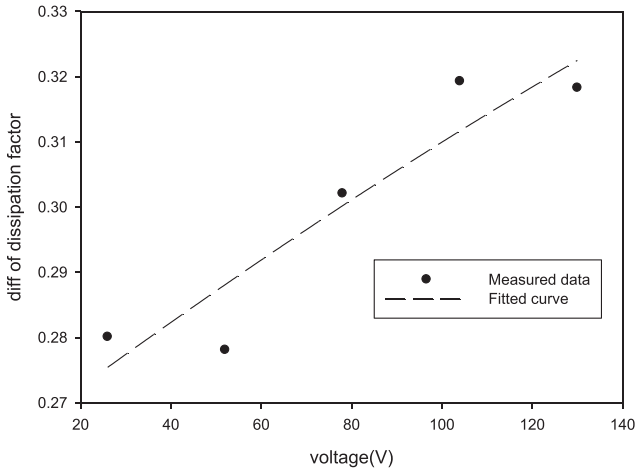


Fig. 13a. Deviation of dissipation factor for different excitation at 50 Hz and 50 °C.

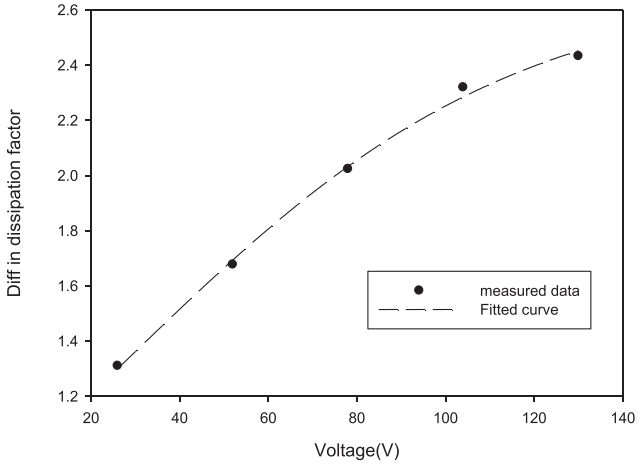


Fig. 13b. Deviation of dissipation factor for different excitation at 50 Hz and 55 °C.

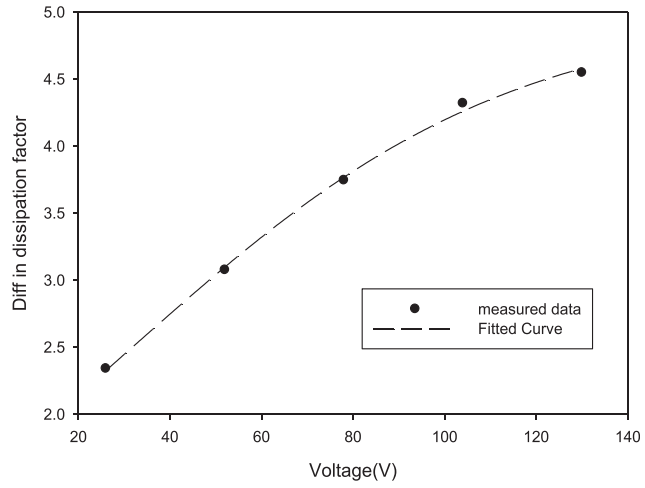


Fig. 13c. Deviation of dissipation factor for different excitation at 50 Hz and 60 °C.

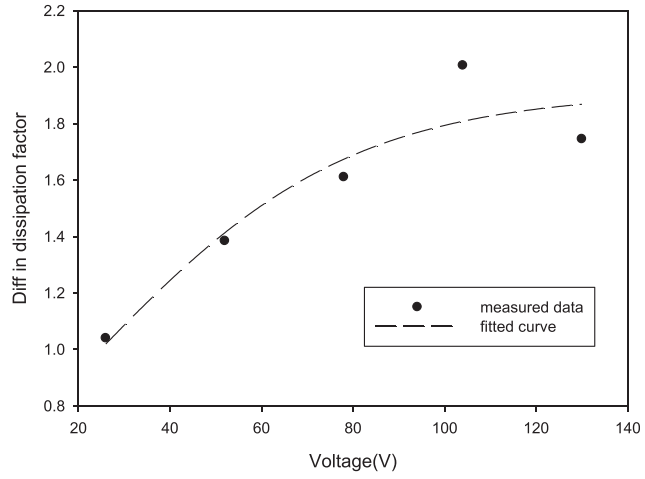


Fig. 13d. Deviation of dissipation factor for different excitation at 50 Hz and 70 °C.

Constants a(T), b(T), and $X_0(T)$ are temperature-dependent function. These values are calculated using different curve fitting shown in Figure 14a, 14b and 14c.

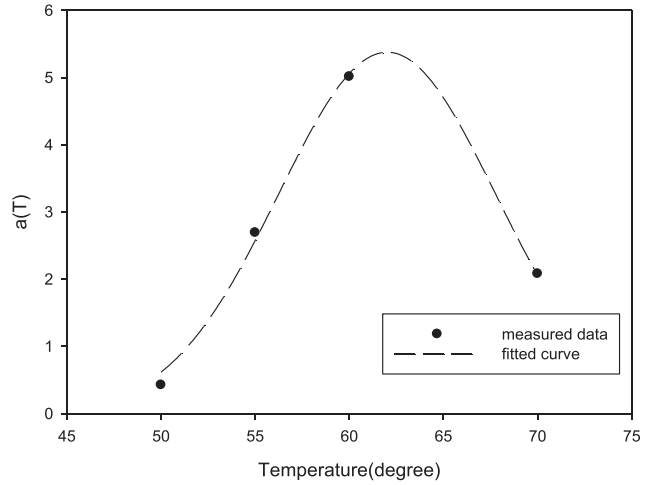


Fig. 14a. Deviation of a(T) for the different temperatures at 50 Hz.

$$a(T) = y_0 + a \times \exp\left(-0.5 \times \left(\frac{T - x_0}{b}\right)^2\right) \quad (6)$$

Constant a(T) is best fitted with Gaussian peak function for three parameters shown in equation (6). The values of constants y_0, a, b and x_0 are −0.6807, 5.9513, 6.5097, and 61.9345.

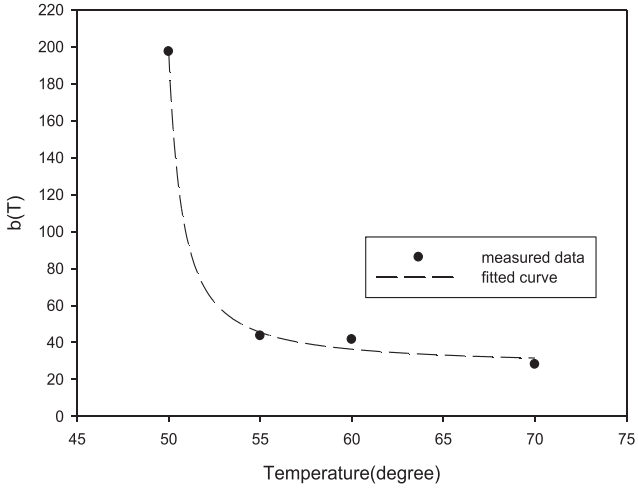


Fig. 14b. Deviation of b(T) for the different temperatures at 50 Hz.

$$b(T) = a \times \exp\left(\frac{b}{T + c}\right) \quad (7)$$

Constant b(T) is best fitted with the decaying exponential function for three parameters shown in equation (7). The values of constants a, b and c are 26.7402, 3.6032, and −48.1974.

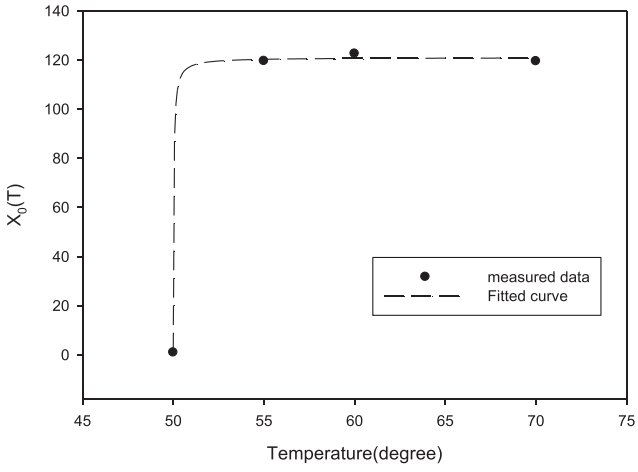


Fig. 14c. Deviation of $X_0(T)$ for the different temperatures at 50 Hz.

$$X_0(T) = a/1 + \exp(x_0 T/b) \quad (8)$$

Constant $X_0(T)$ is best fitted with the hyperbola function for three parameters shown in equation (8). The values of constants a, b and x_0 are 121.4416, −0.0703, and −49.9707.

4.2. 1.2,DecimalComplex,1.2,DecimalComplex, for aged mineral oil at constant frequency (1hz)

The deviation of *tanδ* at different temperatures are calculated from Figure 15a, 15b, 15c and 15d. The value of *Δtanδ (V)* at 50°C is fitted with the sigmoidal function shown in figure 15 (a). Similarly, the value of *Δtanδ (V)* at 55°C, 60°C, and 70°C is also fitted with the sigmoidal function shown in figure 15 (b,c & d). After

the fitting process for different temperatures, the value of $\Delta tan\delta\,(V, T)$ at power frequency (1 Hz) is calculated using a combined sigmoidal curve fitting for all temperatures, which forms equation (9). Equation (9) is a sigmoidal function in which V is a different excitation level provided by the function generator.

If the $tan\delta$ values recorded at the same temperature under different excitation voltages are compared, a non-linear relationship

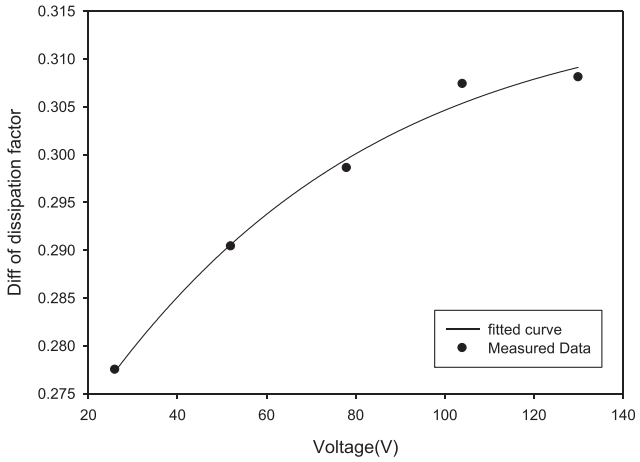


Fig. 15a. Deviation of dissipation factor for different excitation at 1 Hz and 50 °C.

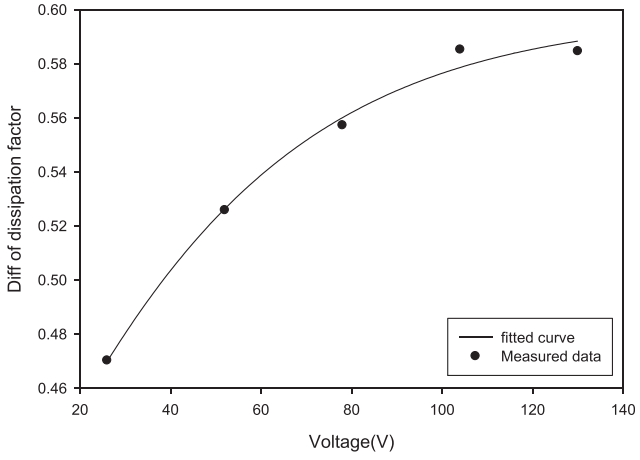


Fig. 15b. Deviation of dissipation factor for different excitation at 1 Hz and 55 °C.

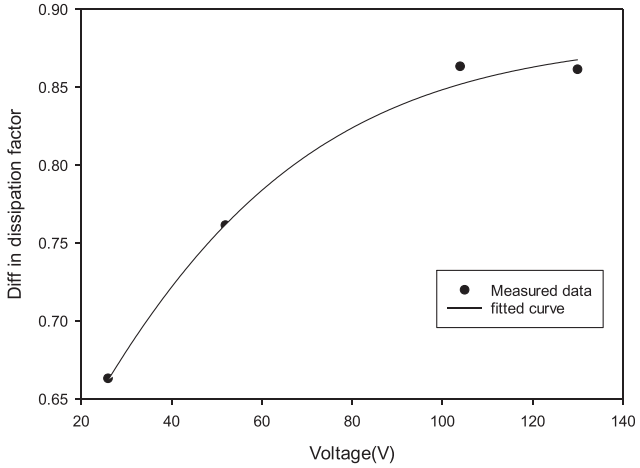


Fig. 15c. Deviation of dissipation factor for different excitation at 1 Hz and 60 °C.

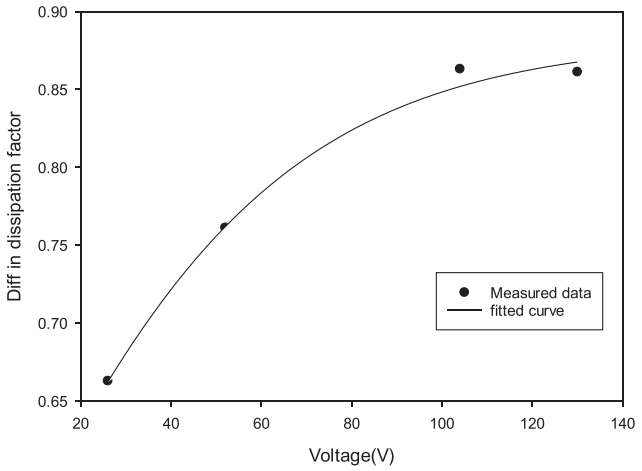


Fig. 15d. Deviation of dissipation factor for different excitation at 1 Hz and 70 °C.

is observed. This is evidenced in Figure 12a, 12b, 12c and 12d. This relationship can be given by the following equation,

$$\Delta \tan\delta(V,T) = \left(\frac{a(T)}{1+\exp\left(X_0(T) - V/b(T)\right)}\right) \tag{9}$$

Constants a(T), b(T), and $X_0(T)$ are temperature-dependent function. These values are calculated using different curve fitting shown in Figure 16a, 16b and 16c.

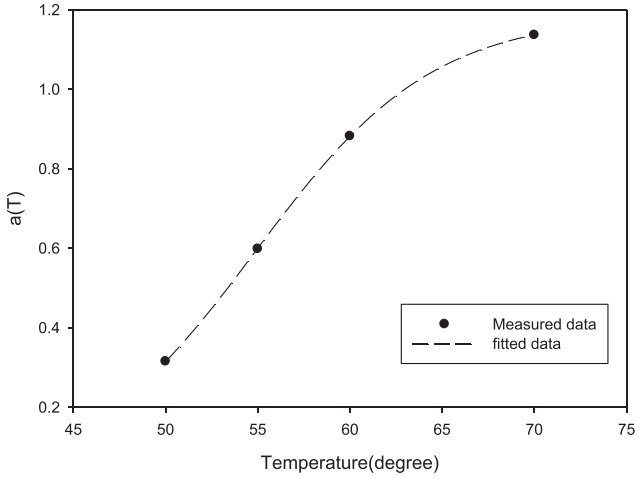


Fig. 16a. Deviation of a(T) for the different temperatures at 1 Hz.

$$a(T) = y_0 + a \times \exp\left(-0.5 \times \left(\frac{T - x_0}{b}\right)^2\right) \tag{10}$$

Constant a(T) is best fitted with Gaussian peak function for three parameters shown in equation (10). The values of constants y_0, a, b and x_0 are −0.6807, 5.9513, 6.5097, and 61.9345.

$$b(T) = a \times \exp\left(\frac{b}{T + c}\right) \tag{11}$$

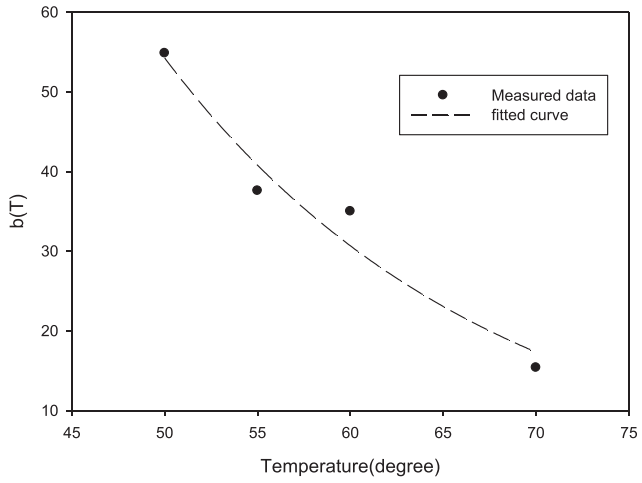


Fig. 16b. Deviation of b(T) for the different temperatures at 1 Hz.

Constant b(T) is best fitted with the decaying exponential function for three parameters shown in equation (11). The values of constants a, b and c are 26.7402, 3.6032, and −48.1974.

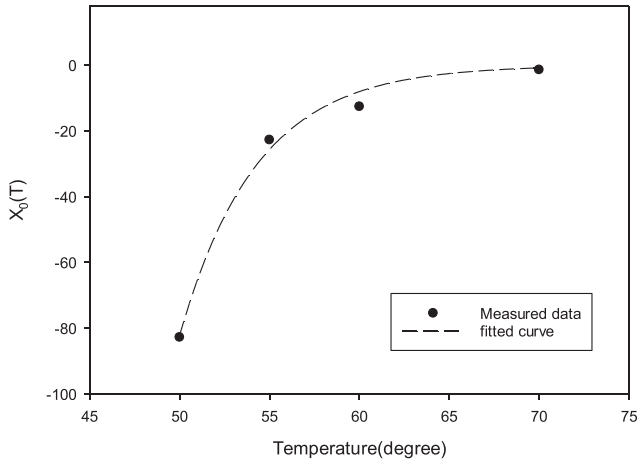


Fig. 16c. Deviation of $X_0(T)$ for the different temperatures at 1 Hz.

$$X_0(T) = a/1 + \exp((x_0 - T)/b) \quad (12)$$

Constant $X_0(T)$ is best fitted with a hyperbola function for three parameters shown in equation (12). The values of constants a, b and x_0 are 121.4416, −0.0703, and −49.9707.

5. Conclusions

This paper discusses dielectric frequency response for silica-based new & aged mineral oil. The most important findings are summarized as follows:

- It is found that non-linearity occurs at different excitation levels. Minimum tanδ was found at a higher value of frequency range with the addition of silica which shows improved dielectric properties of oil-paper insulation.
- As observed in Fig. 9, the value of tanδ at 50 Hz is highly affected by temperature as well as the applied voltage. This may limit the application of silica-based nanofluids in practical transformers, as tanδ at 50 Hz is a reliable insulation performance parameter used by power utilities.
- If the tanδ value is affected by applied excitation and temperature, it may not provide unambiguous information regarding insulation health anymore. The paper presents an equation to evaluate the dissipation factor at different voltages and temperatures on power frequency using only measured dielectric frequency response data at one voltage and one temperature. Equation (5) finds the dissipation factor with an improved dielectric response at power frequency which is dependent on excitation voltage and temperature. Present work attempts to solve this issue. This method is applicable to laboratory samples and power transformers both.

References

[1] Lv YZ, Zhou Y, Li CR, Wang Q, Qi B. Recent progress in nanofluids based on transformer oil: preparation and electrical insulation properties. IEEE Electr Insul Mag 2014;30(5):23–32.
[2] Rafiq M, Lv Y, Li C. A Review on Properties, Opportunities, and Challenges of Transformer Oil-Based Nanofluids. Journal of Nanomaterials 2016;2016:1–23.
[3] Zhou YX, Wang Ys, Tian JH, Sha YC, Jiang XX, Gao SY, Nic Q. Breakdown Characteristics in Transformer Oil Modified by Nanoparticles, in High Voltage Engineering. in High Voltage Engineering 2010;36(5):1155–60.
[4] Choi SUS. Enhancing Thermal Conductivity of Fluids with Nanoparticles. In: Siginer DA, Wang HP, editors. Developments and Applications of Non-Newtonian Flows, 66. New York: ASME; 1995. p. 99–105.
[5] Segal V, Hjorstberg A, Rabinovich A, Nattrass D, Raj K. AC and impulse breakdown strength of a colloidal fluid based on transformer oil and magnetite nanoparticles. In: In IEEE International Symposium on Electrical Insulation. p. 619–22.
[6] Segal V, Rabinovich A, Nattrass D, Raj K, Nunes A. Experimental study of magnetic colloidal fluids behaviour in power transformer. J Magn Magn Mater 2000;215:513–5.
[7] Dong M, Shen LP, Wang H, Wang HB, Miao J. Investigation on the Electrical Conductivity of Transformer Oil-Based AlN Nanofluid. Journal of Nanomaterials 2013:1–7.
[8] Rafiq M, Shafique M, Azam A, Ateeq M. Transformer oil-based nanofluid: The application of nanomaterials on thermal, electrical and physicochemical properties of liquid insulation-A review. Ain Shams Eng J 2021;12(1):555–76.
[9] Jin H, Andritsch T, Tsekmes IA, Kochetov R, Morshuis PHF, Smith JJ. Properties of Mineral Oil-based Silica Nanofluids. IEEE Trans Dielectr Electr Insul 2014;21(3):1100–8.
[10] A. Baral and a. S. Chakravorti, Condition Assessment of Cellulosic Part in Power Transformer Insulation using Transfer Function Zero of Modified Debye Model, IEEE Transactions on Dielectrics and Electrical Insulation, 21 (5) (2014) 2028-2036.
[11] Rafiq M, Lv Y, Li C, Yi K. Effect of different Nanoparticle Types on Breakdown Strength of Transformer Oil. the IEEE Conference 2016.
[12] Liu J, Zhou L, Wu G, Zhao Y, Liu P, Peng Q. Dielectric Frequency Response of Oil-Paper Composite Insulation Modified by Nanoparticles. IEEE Trans Dielectr Electr Insul 2012;19(2):510–20.
[13] Zaengl WS. Applications of Dielectric Spectroscopy in Time and Frequency Domain for HV Power Equipment. IEEE Electr Insul Mag 2003;19(6):9–22.
[14] Mansour D-E, Shaalam EM, Ward SA, El Dein AZ, Karaman HS, Ahmed HM. Multiple nanoparticles for improvement of thermal and dielectric properties of oil nanofluids. IET Sci Meas Technol 2019;13(7):968–74.
[15] Du Y, Lv Y, Li C, Chen M, Zhou J, Li X, et al. Effect of Electron Shallow Trap on Breakdown Performance of Transformer Oil-based Nanofluids. J Appl Phys 2011;110(10):104104.
[16] Pradhan AK, Chatterjee B, Chakravorti S. Estimation of Paper Moisture Content based on Dielectric Dissipation Factor of Oil-paper Insulation under Non-sinusoidal Excitations. IEEE Trans Dielectr Electr Insul 2015;22(2):822–30.
[17] Lv Y, wang W, Ma K, Zhou J, Li C, Wang Q. Nanoparticle Effect on Dielectric Breakdown Strength of Transformer Oil-Based Nanofluids.
[18] Prabu A. Nanoparticles as an additive in biodiesel on the working characteristics of a DI diesel engine. Ain Shams Eng J 2018;9:2343–9.
[19] Maneerot S, Pattanadech N, Kittikhuntharadol Y, Nimsanong P, Bunlaksananusorn C, Carotenuto G. Polarization and Conduction Characteristics of Mineral Oil and Natural Ester Mixed with Nanoparticles. Adv Mater Sci Eng 2022;2022:1–12.
[20] Ahmad F, Ali Khan A, Khan Q, Hussain MR. State-of-Art in Nano-Based Dielectric Oil: A Review, IEEE. Access 2019;7:13396–410.
[21] Abolfathi B. Improvement of Mineral Oil Transformer Properties using the Combinations of Nano and Micro Particles. International Journal of Engineering and Technology (IJET) 2020;12(2):189–96.
[22] Fasehullah M, Wang F, Jamil S, Huang J, Haq IU, Han Q. Effect of SiO2 Nanoparticle's Size and Doping Concentration on AC Breakdown Behavior of Insulating Oil-based Nanofluids December 2021;12–15:343–6.
[23] Khaled U, Beroual A. AC dielectric strength of synthetic ester-based Fe3O4, Al2O3 and SiO2 nanofluids—conformity with normal and weibull distributions. IEEE Trans Dielectr Electr Insul 2019;26(2):625–33.
[24] Fal J, Mahian O, Zyła G. Nanofluids in the Service of High Voltage Transformers: Breakdown Properties of Transformer Oils with Nanoparticles, a Review. Energies 2018;11(2942):1–46.

[25] Dutta S, Baral A, Pradhan AK, Chakrovorti S. Effect of measurement temperature on power transformer insulation diagnosis using frequency-domain spectroscopy. IET Sci Meas Technol 2017;11(6):773–9.



Hari Charan Verma obtained his B.Tech degree from Rajiv Gandhi Proudyogiki Vishwavidyalaya (RGPV), Bhopal, India and M.Tech degree from Delhi Technological University (DTU), Delhi, India in the year 2012 and 2015 respectively. He is presently working as an Assistant Professor in the Department of Electrical Engineering, Muzaffarpur Institute of Technology, Muzaffarpur, India. He is also pursuing Ph.D. from Indian Institute of Technology (Indian School of Mines), Dhanbad, India. His research interest includes condition monitoring of transformers.



Arijit Baral (M2010 - SM2016) received his B.Tech degree in Electrical Engineering from West Bengal University of Technology, Kolkata, India in 2007, M.E. degree from Bengal Engineering and Science University, Shibpur, India in 2009 and Ph.D. degree from Jadavpur University, Kolkata, India in 2015. He worked as an INSPIRE Fellow in Department of Electrical Engineering, Jadavpur University, Kolkata, India from 2011 to 2014. In 2015 he received the prestigious POSOCO Power Systems Award (PPSA). He is presently working as an Assistant Professor in the Department of Electrical Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, India. His current field of interest includes condition monitoring of high voltage power equipment and application of soft computing techniques in Power Systems related fields.

Journal of Electronic Materials
https://doi.org/10.1007/s11664-022-09973-1

REVIEW ARTICLE

Effect of Different Counter Electrodes on Power Conversion Efficiency of DSSCs

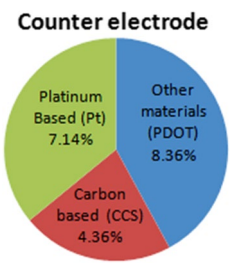
Geetam Richhariya[1] · Anil Kumar[2,3] · Akash Kumar Shukla[4] · K. N. Shukla[5] · Bhim Charan Meikap[6]

Received: 2 April 2022 / Accepted: 15 September 2022
© The Minerals, Metals & Materials Society 2022

Abstract

Among various photovoltaic (PV) devices, dye-sensitized solar cells (DSSCs) represent an outstanding choice due to their low cost, easy fabrication procedures, and eco-friendly nature. The photoanode, sensitizer, electrolyte, and counter electrode are basic elements that contribute to the performance of DSSCs. Counter electrode (CE) materials play a significant role in the DSSC PV performance. Improving cost-effective electrodes for PV applications is vital to meeting potential power demand. As an essential element, CE catalysts play a crucial role in DSSCs. An expensive platinum-based CE can be replaced by a number of economical and extremely stable materials such as conducting polymers, sulfides, oxides, and carbonaceous materials such as N-doped core–shell-based DSSCs, which have reportedly attained power conversion efficiency of 7.89%, higher than platinum-based cells (7.48%). Cr-doped SiC showed better activity for triiodide splitting; however, it also revealed poor action for further splitting of the iodine into mono-iodides as compared with Pt-doped SiC slabs. A graphene nanoplatelet-based CE revealed charge transfer resistance of 3.36 Ω, slightly higher than platinum (1.18 Ω). Consequently, it showed slightly higher oxidation. The present article focuses on various striking replacements for platinum-based CEs for DSSCs that will offer innovative study opportunities for different CE materials in renewable energy fields.

Graphical Abstract



Keywords Counter · electrode · DSSC · platinum · fill factor

✉ Anil Kumar
anilkumar76@dtu.ac.in

[1] Department of Electrical and Electronics Engineering, Oriental Institute of Science and Technology, Bhopal 462022, India

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

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

[4] Department of Electrical Engineering, Ujjain Engineering College, Ujjain 456010, India

[5] Department of Electrical Engineering, LNCT College, Bhopal 462021, India

[6] Department of Chemical Engineering, Indian Institute of Technology, Kharagpur 721302, India

Published online: 20 October 2022

Introduction

The recent focus on sustainable energy development has driven research advancements in low-cost photovoltaic cells. Dye-sensitized solar cells (DSSCs) have emerged from the current revolution in photovoltaics and have garnered significant attention owing to their low cost and simple fabrication.[1] Photoanodes, photosensitizers, counter electrodes (CE), and electrolytes are the main elements of traditional DSSCs. The properties of fabrication materials and charge transfer dynamics (i.e., assortment of electrons at anode, accumulation of light of photoanode, and rapid decrease of an electrolyte at cathode side) affect the performance of the DSSC.[2] The key element of a dye solar cell is the counter electrode, which plays a vital role in cell power conversion efficiency (PCE).[3,4] A DSSC's overall efficiency relies on the properties of these materials and the kinetics of charge transport within the cell.[5–7] Platinum is the most effective counter electrode catalytic material and also has excellent conductivity and higher iodine corrosion stability in the electrolyte for DSSC applications.[8] However, lack of availability resources, high cost, and oxidation of liquid electrolytes restrict the large-scale processing of DSSCs.[9] Therefore, alternatives have also been explored, such as carbon allotropes, conductive polymers and their composites, transition metals, and sulfide materials. The performance of a black blend of graphitic carbon as a counter electrode material for DSSC was 6.7%.[10] Since then, different types of carbon have been studied as counter electrode materials for DSSCs, including mesoporous carbon, activated carbon, fullerene, carbon nanotubes, graphite, and graphene. Graphene has garnered increasing interest among all carbon polymorphs because of its properties such as strong electrochemical behavior, high electrical and thermal conductivity, and substantial transparency across the entire solar spectrum.[11] DSSCs fabricated with graphene-based counter electrodes need large amounts of graphene in bulk quantities. Liquid-phase shear exfoliation of graphite into graphene is the most cost-effective means of bulk production among the different methods available for producing graphene.[12] Several reports of graphene derived from shear exfoliation in liquid dispersions have been published. The primary aim of this review is to highlight the various counter electrode materials of DSSCs that affect their performance. Various concentrations of graphene oxide sheets (GS) were synthesized with platinum (Pt) precursor, and the effects of the GS-Pt compound were studied for application as counter electrode. GS-Pt composite-based DSSCs with 0.01 wt.% GS have shown energy conversion efficiency up to 6.6%.

Additionally, GS enhances Pt crystallinity and surface area, sparing catalytically active sites for reduction of triiodide.[13] The effect of vitreous carbon (VC), fluorine-doped tin oxide (FTO), and indium tin oxide (ITO) on the electrocatalytic properties of Cu_2ZnSnS_4 (CZTS) and $Cu_2ZnSnSe_4$ (CZTSe) nanoparticles was studied. CZTS–ITO demonstrated outstanding performance, with PCE of 3.62%. Hence, the type of substrate used influenced the overall solar cell performance. CZTS showed better electrocatalyic properties than CZTSe, which could be because of SeOx species on the CZTSe surface.[14] The PCE of DSSCs strongly depends on the counter electrode. Carbon, graphite, and gold were employed as counter electrodes to fabricate hibiscus-based DSSCs. It was observed that the DSSC using gold as counter electrode had the highest PCE of 0.23%. This could be because more electrons accumulated at the TiO_2 side, resulting in improved cell current density.[15]

Effect of Different Counter Electrode Materials on Power Conversion Efficiency

The efficiency of counter electrodes is affected by various parameters. Counter electrodes are broadly classified into three categories, namely carbon-based, platinum-based, and other material (metals sulfides)-based, as shown in Fig. 1.

Carbon-Based Counter Electrodes

Hexagonal palladium-carbon nanocubes (H–Pd/CNCs) were prepared using a simple single-step chemical synthesis procedure, and the structured materials were then used as the counter electrode to replace the use of platinum-based electrodes for the DSSC. Brunauer–Emmett–Teller (BET) analysis showed that the H–Pd/CNC sample had a large surface area (568.8 $m^2\ g^{-1}$), with average pore size of ~3 nm.[16] PCE of 4.1% was obtained using H–Pd/C NCs as counter electrode. Natural graphite was used to prepare graphene nanoplatelets (GNPs) using a simple and efficient liquid-phase high-shear exfoliation method.[17] GNPs were used as the counter electrode material for DSSCs; under the standard conditions, photoconversion efficiency was 6.23%.[18] Few-layer graphene (FLG) composite conductive inks were synthesized by liquid-phase high shear exfoliation. FLG inks were prepared with multi-walled carbon nanotubes (MWCNT) and carbon black (CB) in polyimide matrix. The prepared ink formulations were appropriate for high-temperature processing at 500°C, allowing graphene-based synthesized conducting films on steel. The cyclic voltammetry (CV) of FLG composite conductive ink-coated steel showed encouraging catalytic activity for redox reaction of I^-/I_3^- redox mediator, as shown in Fig. 2.[19]

A metal-free conductive catalyst was developed based on N-doped core–shell (N-CCS) carbon spheres through rich, dynamic sites. Moreover, the fully interconnected mesoporous basal planes of carbon spheres were activated

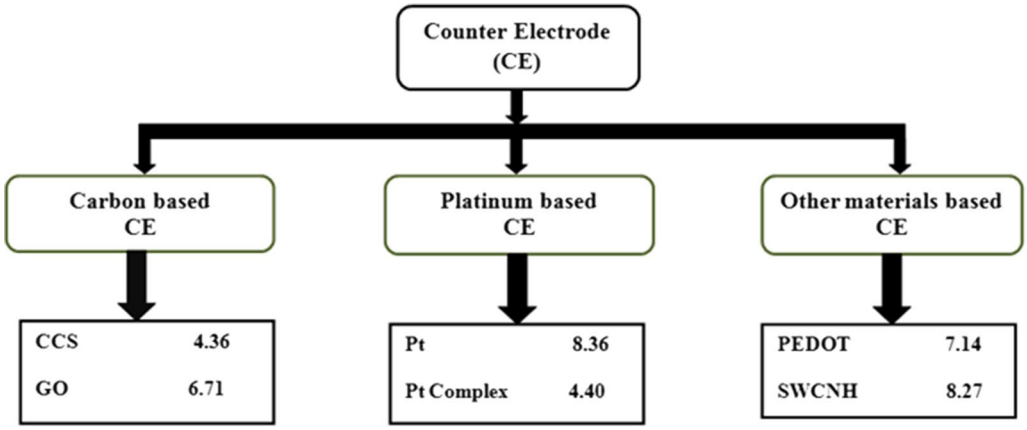


Fig. 1 Overview of different counter electrode materials for DSSCs.

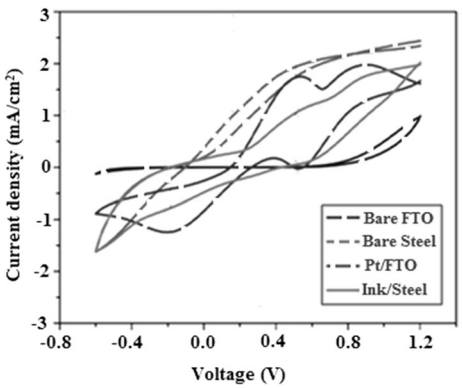


Fig. 2 Cyclic voltammetry of developed ink-coated steel, Pt-coated FTO, bare steel, and bare FTO-based counter electrodes[19] [Reprinted from Solar Energy, 169, Siva Sankar Nemala, Purnendu Kartikay, Rahul Kumar Agrawal, Parag Bhargava, Sudhanshu Mallick, Sivasambu Bohm, Few layers graphene based conductive composite inks for Pt free stainless steel counter electrodes for DSSC, 67–74, copyright (2018), with permission from Elsevier].

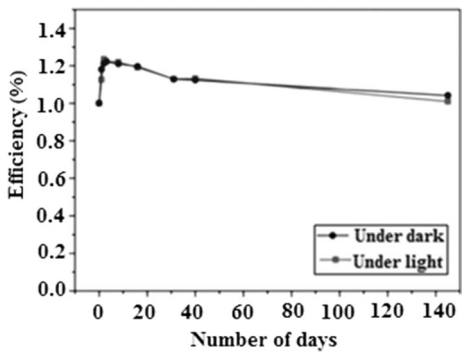


Fig. 3 Stability of DSSCs under constant illumination[18] [Reprinted from Electrochimica Acta, 285, Siva Sankar Nemala, Karanveer S. Aneja, Parag Bhargava, H.L. Mallika Bohm, Sudhanshu Mallick, Sivasambu Bohm, Novel high-pressure airless spray exfoliation method for graphene nanoplatelets as a stable counter electrode in DSSC, 86–93, copyright (2018), with permission from Elsevier].

Table I Photovoltaic parameters of different CEs[20]

Counter electrode	V_{oc} (V)	J_{sc} (mA cm^{-2})	FF	PCE (%)
Pt	0.750	13.7	0.54	7.48
CCS	0.750	12.4	0.47	4.36
N-CCS	0.745	15.3	0.69	7.89

due to doping of N atoms. The N-CCS-based DSSC attained PCE of 7.89%, higher than platinum-based cells (7.48%), as shown in Table I.[20]

For the DSSC with a counter electrode synthesized by reduced graphene oxide (rGO) using sodium 1,4-bis (neopentyloxy)-3-(neopentyloxycarbonyl)1,4-dioxobutane-2-sulphonate (TC_{14}) as a hyperbranched surfactant, energy conversion efficiency of 0.0266% was obtained using this novel CE-based DSSC. High-pressure airless spray was adopted to produce large amounts of GNPs. The GNPs were used as a counter electrode in the DSSC. Good stability was observed over 3000 h under constant illumination without any noteworthy reduction in efficiency, as shown in Fig. 3.[21]

Reduced graphene oxide (SSrGO) was created by a Xe irradiation source. A suspension of SSrGO was spread onto the glass substrate, accompanied by spray-coating of single-walled carbon nanohorns (SWCNH) suspension. The DSSC test cell showed PCE of 8.27% using SWCNH/SSrGO, as shown in Table II.[22]

A DSSC was prepared using a structured configuration (n-MWCNT-TiO_2/N_3/MWCNT). Here, Pt was replaced by inexpensive MWCNT that enables lower-cost DSSC fabrication. Figure 4 shows the absorbance spectrum of the dye, which demonstrates that the presence of MWCNT in the nMWCNT-TiO_2 nanocomposite facilitates the excitation of

Table II Photovoltaic performance of DSSC[22]

Counter electrodes	V_{oc} (V)	J_{sc} (mA/cm^2)	FF	η (%)
GO	0.70	14.53	0.62	6.71
SWCNH/SSrGO	0.73	17.22	0.65	8.27
Pt	0.72	18.55	0.62	8.36

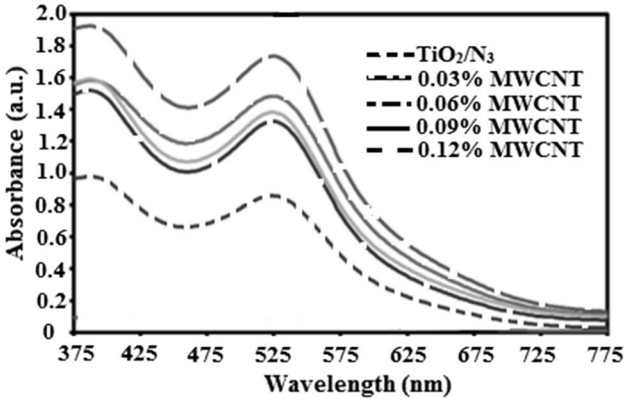


Fig. 4 UV–Vis absorption spectra of N3 dye in TiO_2/N_3 and n-MWCNT-TiO_2/N_3,[23] [Reprinted from Solar Energy, 188, M. Younas, M.A. Gondal, M.A. Dastageer, K. Harrabi, Efficient and cost-effective dye-sensitized solar cells using MWCNT TiO_2 nanocomposite as photoanode and MWCNT as Pt-free counter electrode, 1178–1188, copyright 2019, with permission from Elsevier].

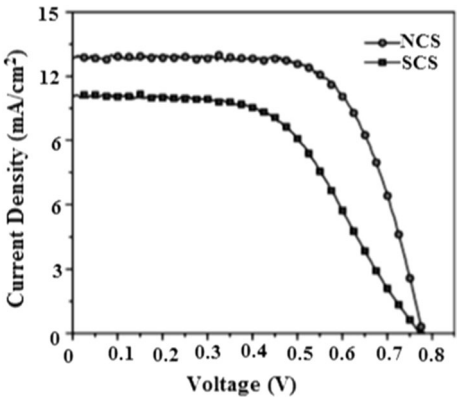


Fig. 5 Power conversion efficiency of DSSC using solid and nitrogen-doped carbon sphere[24] [Reprinted from Electrochimica Acta, 334, Hongrui Wang, Jing Gao, Jingqian Zhu, Jing-Yuan Ma, Huawei Zhou, Jun Xiao, Mingxing Wu, Design bifunctional nitrogen doped flexible carbon sphere electrode for dye-sensitized solar cell and supercapacitor, 135582–0, copyright 2020, with permission from Elsevier].

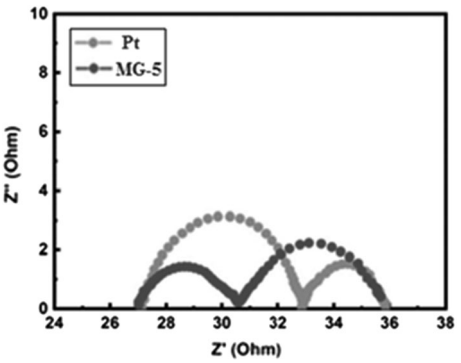


Fig. 6 EIS of symmetric cells of Pt- and MG-5-based counter electrodes[26] [Reprinted from Journal of Electroanalytical Chemistry, 847, R. Senthilkumar, Murali Balu, S. Ramakrishnan, Praveen C. Ramamurthy, Sudip K. Batabyal, Duraisamy Kumaresan, Nikhil K. Kothurkar, Molybdenum disulfide/reduced graphene oxide hybrids with enhanced electrocatalytic activity: An efficient counter electrode for dye-sensitized solar cells, 113236–0, copyright (2019), with permission from Elsevier].

electrons into the conduction band of MWCNT-TiO_2 nanocomposite semiconductors that are initiated by dye.[23]

The utilization of liquid electrolytes in DSSCs is a considerable issue limiting their commercialization. Still, the initiative of leakage of electrolytes intensely alters the reliability of DSSCs, which are likely to fail for an extended time when used outdoors. The identity-charged integrated electronics that connect the energy conversion unit of the DSSC and the energy storage unit of the supercapacitor require a more efficient double-featured electrode material. Nitrogen-doped carbon spheres (NCS) were included to stimulate electrode activity. Compared to traditional solid carbon spheres, PCE of 8.64% was obtained with NCS, as shown in Fig. 5.[24]

The immaculate perfect and imperfect graphene nanosheets (GNSs) were created and used as substrates for catalytic activity against the T_2 to T^- reduction process. Exothermic adsorption reaction discovered that virgin GNSs could achieve high reactivity to T_2/T^-.[25] The development of low-cost, metal-free electrodes for use in solar cells is critical for achieving growing energy demands. Molybdenum disulfide/reduced graphene oxide (MoS_2-rGO) hybrids were great achievements for counter electrode materials for DSSCs. Electrochemical impedance (Fig. 6) and cyclic voltammetry (Fig. 7) studies have revealed that the integration of reduced graphene oxide in the MoS_2 nanostructures significantly enhanced the electrocatalytic activity of CEs and decreased the charge transfer resistance.[26]

Platinum-Based Counter Electrodes

The large Pt nanoparticles decorated with reduced graphene oxide (L-Pt/rGO) was used as the counter electrode for DSSCs. L-Pt/rGO was prepared at room temperature by irradiating γ-rays on the homogenous aqueous solution of GO with 10 wt.% isopropyl alcohol aqueous solution consisting of Pt precursor. Figure 8 reveals the photocurrent

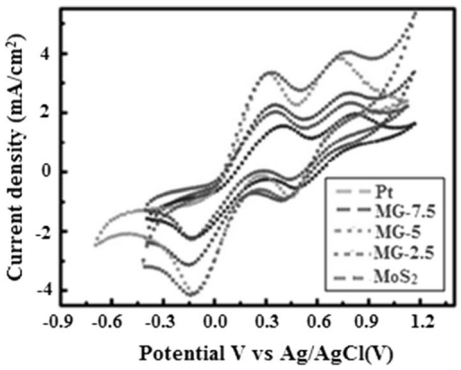


Fig. 7 Cyclic voltammograms of Pt, MG-7.5, MG-5, MG-2.5, and MoS_2 -based counter electrodes[26] [Reprinted from Journal of Electroanalytical Chemistry, 847, R. Senthilkumar, Murali Balu, S. Ramakrishnan, Praveen C. Ramamurthy, Sudip K. Batabyal, Duraisamy Kumaresan, Nikhil K. Kothurkar, Molybdenum disulfide/reduced graphene oxide hybrids with enhanced electrocatalytic activity: An efficient counter electrode for dye-sensitized solar cells, 113236–0, copyright (2019), with permission from Elsevier].

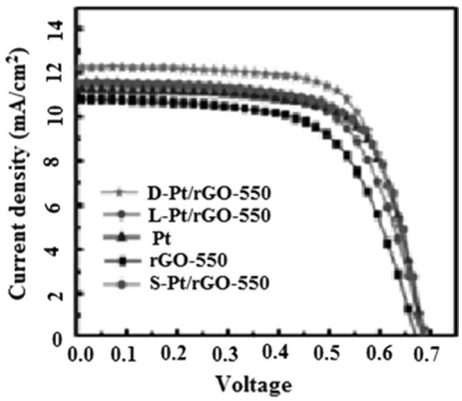


Fig. 8 Power conversion efficiency for various DSSCs[27] [Reprinted from Journal of Industrial and Engineering Chemistry, 65, Hyo-Jin Ahn, Jung-Soo Lee, Hyo-Sub Kim, In-Tae Hwang, Ji-Hyun Hong, Junhwa Shin, Chan-Hee Jung, Fabrication of large Pt nanoparticles-decorated rGO counter electrode for highly efficient DSSCs, 318–324, copyright (2018), with permission from Elsevier].

density–voltage curves for DSSCs with the rGO-550, L-Pt/rGO-550, Pt, S-Pt/rGO-550 and D-Pt/rGO-550 counter electrode.[16,27]

The efficiency of a tungsten oxide/titanium oxide nanocomposites (nWO_3-TiO_2) solar cell with three dissimilar WO_3-TiO_2 [1% WO_3-TiO_2/N719/MWCNT] was improved by around 40% compared to a normal TiO_2/dye/Pt solar cell as revealed in Fig. 9.[28]

A stability test was described for cobalt complex-based DSSCs using platinum or poly(3,4-ethylene dioxythiophene) as counter electrode catalyst. The stability of the device was affected by Co(III) concentration and the counter electrode catalyst as given in Table III. The degradation process was

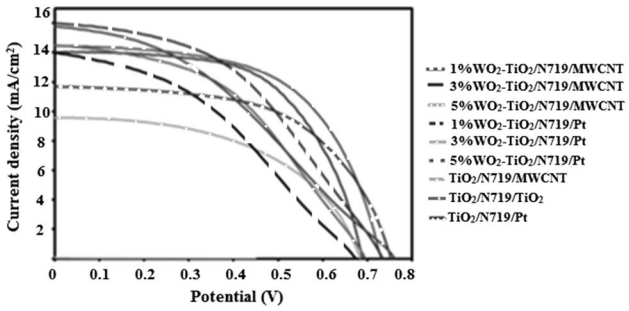


Fig. 9 Power conversion efficiency of DSSCs using nWO_3-TiO_2 electrode[28] [Reprinted from Ceramics International, 45, M. Younas, M.A. Gondal, M.A. Dastageer, U. Baig, Fabrication of cost effective and efficient dye sensitized solar cells with WO_3-TiO_2 nanocomposites as photoanode and MWCNT as Pt-free counter electrode, 936–947, copyright (2019), with permission from Elsevier].

slowed due to a higher concentration of charge carriers. However, increasing the concentration did not improve the degradation resistance.[29]

The efficiency of electrodes produced by aerosol-assisted chemical vapor deposition (AACVD) is found to be better than conventional Pt-counter electrodes made using the doctor blade method. Table IV shows that the performance of a counter electrode produced using the AACVD procedure was superior to that of a platinum counter electrode prepared by the commercial doctor blade technique. Higher open-circuit voltage was observed in the case of Pt complex 2 due to a decreased probability of recombination at the working electrode/electrolyte interface.[30]

The performance of DSSCs was enhanced using a composite counter electrode prepared from 3.0 ml colloidal SnO_2 solution and 0.05 g of graphite powder. Table V shows the improvement in efficiency due to the deeply porous nanostructure of the graphite/SnO_2 composite, providing more reaction sites for triiodide ion reduction.[31]

Sample A (as-synthesized) and sample B (annealed at 100°C) of ZnO@WO_3 core/shell nanoparticles were considered as counter electrodes with a methyl orange-based DSSC. Sample B demonstrated higher PCE of 5.73% (Table VI). This may be due to lower surface resistivity, better electrical conductivity, and ZnO@WO_3 core/shell nanoparticles. The innovative materials benefit from the advantages of two materials in ZnO@WO_3 nanocomposites, and therefore this complex material combines their advantages to achieve the vital behavior. However, sample A shows low light absorption efficiency, leading to poor electron generation.[32]

Table III Stability and performance parameters of PDOT- and Pt-based DSSCs[29]

Counter electrodes	Hours (h)	V_{oc} (V)	J_{sc} (mA/cm^2)	FF	η (%)
PDOT	0 h	0.723 ± 0.11	10.6 ± 0.4	0.68 ± 0.5	5.2 ± 0.4
	1000 h	0.581 ± 0.12	5.2 ± 0.4	0.44 ± 0.1	1.3 ± 0.1
	2000 h	0.540 ± 0.12	4.4 ± 1.7	0.51 ± 0.9	1.6 ± 0.6
Pt	0 h	0.779 ± 0.60	11.4 ± 0.2	0.68 ± 0.3	6.0 ± 0.2
	1000 h	0.667 ± 0.11	10.8 ± 0.4	0.34 ± 0.3	2.9 ± 0.2
	2000 h	0.642 ± 0.12	7.9 ± 0.3	0.37 ± 0.2	1.9 ± 0.1

Table IV Performance of Pt complex-based cells[30]

Counter electrode	V_{oc} (V)	J_{sc} (mA/cm^2)	FF	η (%)
Pt-complex 1	0.751	11.40	0.432	3.69
Pt-complex 2	0.765	13.34	0.423	4.32
Pt-commercial	0.771	14.62	0.389	4.40

Table V Performance of composite counter electrode-based cells[31]

Counter electrodes	V_{oc} (V)	J_{sc} (mA/cm^2)	FF	η (%)
Pt	0.803	16.43	0.70	8.12
Graphite/SnO_2	0.693	18.47	0.52	6.71

Table VI Performance parameters of Zn@WO_3 core–shell nanoparticle-based DSSC[32]

Counter electrode	V_{oc} (V)	J_{sc} (mA cm^{-2})	FF	PCE (%)
Pt	0.62	11.4	0.52	4.51
Sample A	0.82	11.7	0.57	4.55
Sample B	0.89	12.1	0.60	5.73

Other Materials for Counter Electrodes

Metal sulfide nanoparticles consisting of PbS, Ag_2S, CuS, CdS, and ZnS are used to replace the costlier platinum counter electrode against the regeneration of organic sulfide redox couples. Figure 10 shows the SEM images of the prepared metal sulfides. Generally, the SEM images reveal a definite degree of aggregation due to the adhesion of the metal sulfide particles. Morphological analysis shows differences in the particle size among the metal sulfides. PBS and CuS show similar morphology, while PbS particles are larger than CuS. The diameters of PbS and CuS are 20–100 and 20–50 nm, respectively.[33]

Replacing platinum with Cu_2ZnSnS_4 (CZTS) nanoparticles for the counter electrode material and TiO_2 nanobelts as the anode for DSSC showed PCE of 5.76%. Figure 11 shows the X-ray diffraction (XRD) patterns of the CZTS nanoparticles, which reveal three dominant peaks at (112), (220), and (312) ascribed to the CZTS kesterite phase.[34]

Quasi-solid electrolytes with poly(aniline) (PANI) counter electrode showed superior results as compared with platinum as counter electrode, as illustrated in Table VII.[35]

A suitable counter electrode for DSSCs must have excellent electrocatalytic characteristics and must be cost-effective. The PCE of DSSCs with the ideal PEDOT-Ni_2P-3 electrode and PEDOT-Co_2P-3 electrode was 7.14% and 6.85%, respectively, as described in Table VIII.[36]

Because of its low cost, PANI is a compelling alternative to platinum for DSSCs. Deposition of a PANI layer on a fluorine-doped tin oxide glass surface, on the other hand, is extremely difficult. Photoconversion efficiency of DSSC with Gr/PANI and with a Pt counter electrode was 3.58% and 3.97%, respectively, as revealed in Fig. 12.[37]

Low-cost lead sulfide (PbS) was used as counter electrode, and lead xanthate was employed as a precursor material to synthesize PbS. However, the PCE using lead sulfide (4.71%) was slightly inferior to that of the platinum-based DSSC (5.34%), as shown in Table IX.[38]

Tantalum-formed bimetallic oxides ($FeTa_2O_6$, $CoTa_2O_6$, and $CuTa_{10}O_{26}$) and bio-based spherical carbon (SC)-supported composites ($FeTa_2O_6$/SC, $CoTa_2O_6$/SC, and $CuTa_{10}O_{26}$/SC) were effectively prepared using a hydrothermal/annealing method, and were employed as CEs catalysts for DSSCs. Aloe peel waste was used to obtain bio-based spherical carbons through hydrothermal carbonization. $FeTa_2O_6$/SC, $CoTa_2O_6$/SC, and $CuTa_{10}O_{26}$/SC CE catalysts showed outstanding electrocatalytic activity as well as good electrocatalytic stability for I_3 reduction, revealing PCE of 6.85%, 6.79%, and 6.53% respectively as shown in Fig. 13.[39]

The enhancement in the surface area provided by zinc oxide nanowires (ZnO NWs) by changing the quantity of deposited platinum arrays showed an increase in the peak current density and the reaction rate for I_3^- reduction. The efficiency of the developed DSSCs was 12% and 14% greater than comparable DSSCs with evaporated and electrodeposited electrodes, respectively, as revealed in Fig. 14.[40]

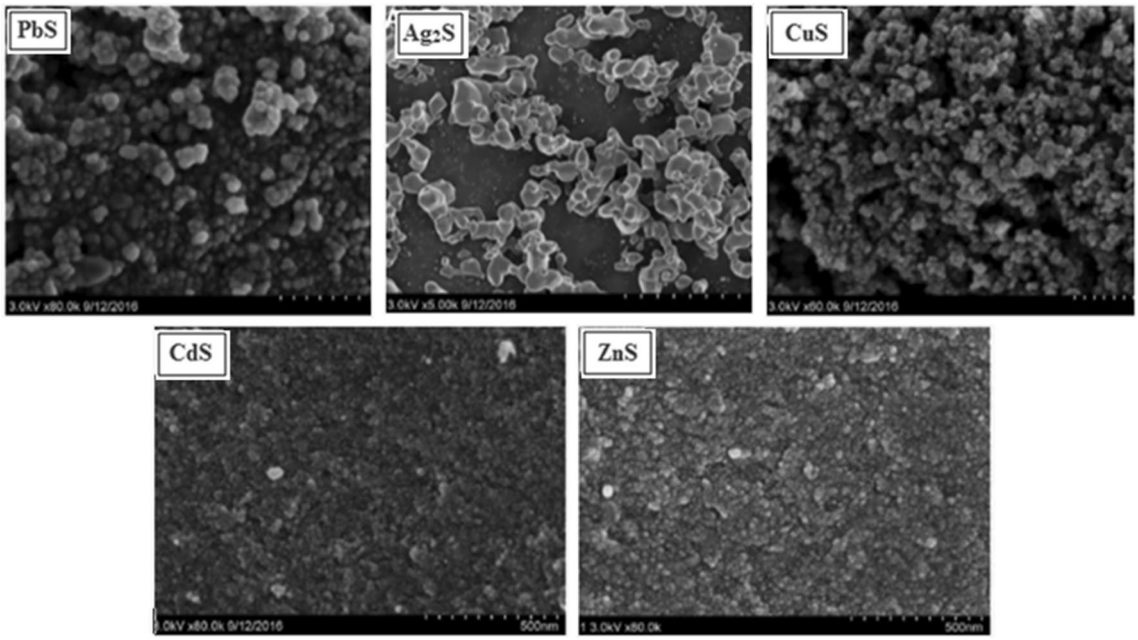


Fig. 10 SEM images for developed metal sulfide powders[33] [Reprinted from Materials Today Energy, 8, Qianji Han, Zaixin Hu, Hongrui Wang, Yuxin Sun, Jing Zhang, Lin Gao, Mingxing Wu, High performance metal sulfide counter electrodes for organic sulfide redox couple in dye-sensitized solar cells, 1–7, copyright (2018), with permission from Elsevier].

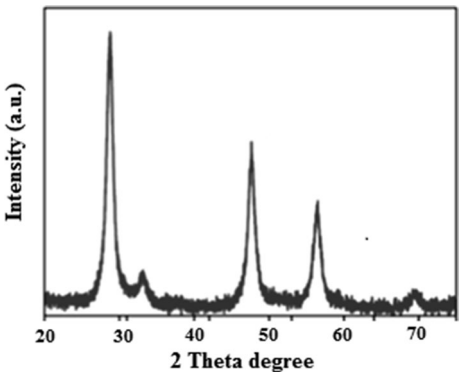


Fig. 11 XRD pattern of the CZTS nanoparticles[34] [Reprinted from Materials Today: Proceedings, 5, Siva Sankar Nemala, Krishnaiah Mokurala, Parag Bhargava, Sudhanshu Mallick, Titania Nanobelts as a Scattering Layer with Cu_2nSnS_4 as a Counter Electrode for DSSC with Improved Efficiency, 23,351–23,357, copyright (2018), with permission from Elsevier].

Table VII Performance parameters of DSSCs using Pt and PANI CEs[35]

Counter electrode	V_{oc} (V)	J_{sc} (mA cm^{-2})	FF	PCE (%)
PANI	0.66	16.2	0.58	6.3
Pt	0.69	14.5	0.66	6.6

Table VIII Performance parameters of PEDOT-Ni_2P-3- and PEDOT-Co_2P-3-based DSSCs[36]

Counter electrode material	V_{oc} (V)	J_{sc} (mA cm^{-2})	FF	PCE (%)
PEDOT-Ni_2P-3	0.75	13.91	0.68	7.14
PEDOT-Co_2P-3	0.75	13.61	0.68	6.85

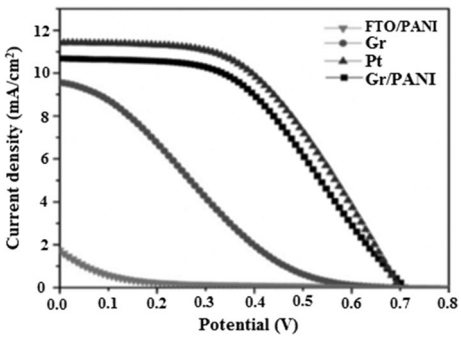


Fig. 12 Power conversion efficiency of DSSCs using FTO/PANI, Gr, Pt, Gr/PANI-based counter electrodes[37] [Reprinted from Diamond and Related Materials, 94, Muhammad Umair Shahid, Norani Muti Mohamed, Ali Samer Muhsan, Robabeh Bashiri, Adel Eskandar Shamsudin, Siti Nur Azella Zaine, Few-layer graphene supported polyaniline (PANI) film as a transparent counter electrode for dye-sensitized solar cells, 242–251, copyright (2019), with permission from Elsevier].

Table IX Photovoltaic performance of DSSCs[38]

Counter electrode	V_{oc} (V)	J_{sc} (mA/cm^2)	FF	PCE (%)
44 mM of $Pb(S_2COEt)_2$	7.32	14.283	0.48	4.71
66 mM of $Pb(S_2COEt)_2$	7.30	13.981	0.45	4.43
Pt	7.29	17.87	0.44	5.34

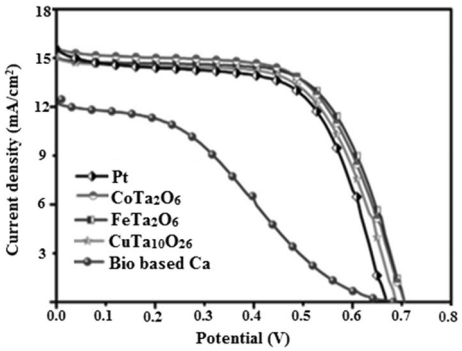


Fig. 13 Power conversion efficiency of DSSCs using $FeTa_2O_6$/SC, $CoTa_2O_6$/SC, and $CuTa_{10}O_{26}$/SC-based counter electrodes[39] [Reprinted from Electrochimica Acta, 309, Sining Yun, Xiao Zhou, Yangliang Zhang, Chen Wang, Yuzhi Hou, Tantalum-based bimetallic oxides deposited on spherical carbon of biological origin for use as counter electrodes in dye sensitized solar cells, 371–381, copyright (2019), with permission from Elsevier].

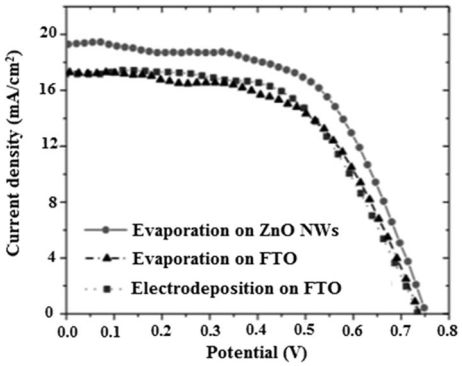


Fig. 14 Power conversion efficiency of DSSCs prepared with the various counter electrodes[40] [Reprinted from Journal of Electroanalytical Chemistry, 835, George Syrrokostas, Katerina Govatsi, George Leftheriotis, Spyros N. Yannopoulos, Platinum decorated zinc oxide nanowires as an efficient counter electrode for dye sensitized solar cells, 86–95, copyright (2019), with permission from Elsevier].

Micro/nanoscale binary (CoS, NiS, and ZnS) and ternary ($CoNi_2S_4$ and $Zn_{0.76}Co_{0.24}S$) metal sulfides were developed on FTO substrates using an innovative, one-step hydrothermal method for economic DSSC counter electrodes. XRD analysis (Fig. 15) showed the configuration of phase pure binary CoS, NiS, ZnS, and ternary $CoNi_2S_4$ and $Zn_{0.76}Co_{0.24}S$ compounds by the careful control of synthesis conditions.[41] The performance parameters of various DSSCs using different counter electrodes are summarized in Table X.

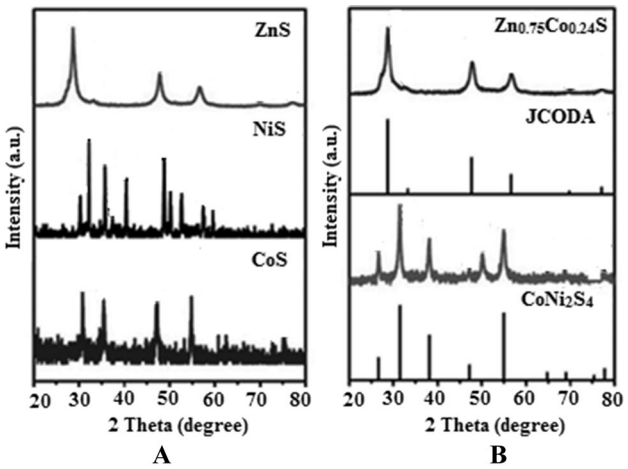


Fig. 15 XRD patterns of (a) binary- and (b) ternary-based counter electrodes[41] [Reprinted from Solar Energy, 193, K. Subalakshmi, K. Ashok Kumar, Onuoha Prayer Paul, S. Saraswathy, A. Pandurangan, J. Senthilselvan, Platinum-free metal sulfide counter electrodes for DSSC applications: Structural, electrochemical and power conversion efficiency analyses, 507–508, copyright (2019), with permission from Elsevier].

Effect of Temperature on the Performance of DSSCs

The most significant external factor influencing the photovoltaic performance of DSSCs is temperature. The temperature of the cell affects the overall performance parameters of the DSSC. The performance of solar cells sensitized by organic dye was investigated at different outdoor temperatures. It was observed that the efficiency of the cell declines when the temperature increases. As mentioned in Table XI, conversion efficiency of 2.58% was recorded at 25°C under air mass (AM) 1.5 G illumination at 100 mW/cm^2. The EIS study revealed a reduction in charge recombination resistance with increasing cell temperature, mainly because of a decline in electron lifetime.[42]

The PCE of a Ru (II) polypyridyl dye (N719)-sensitized ZnO-based photoanode solar cell was investigated at different temperatures. At 40°C, maximum conversion efficiency was recorded as 2.25% under 1 sun 1.5 am, as shown in Table XII. However, the PCE consistently decreased at increasing temperatures, most likely due to reduced open-circuit voltage and fill factor.[43]

Mo and Ti wires were employed as electrodes at low temperature under a water direct-current (DC) plasma technique.

Table X Summary of the counter electrode materials

Counter electrode material	V_{oc} (V)	J_{sc} (mA cm^{-2})	FF	PCE (%)	References
CCS	0.750	12.4	0.47	4.36	20
N-CCS	0.745	15.3	0.69	7.89	
GO	0.70	14.53	0.62	6.71	22
SWCNH/SSrGO	0.73	17.22	0.65	8.27	
Pt	0.72	18.55	0.62	8.36	
PDOT	0.723 ± 0.11	10.6 ± 0.4	0.68 ± 0.5	5.2 ± 0.4	29
Pt-Complex 1	0.751	11.40	0.432	3.69	30
Pt-Complex 2	0.765	13.34	0.423	4.32	
Pt-Complex 3	0.771	14.62	0.389	4.40	
Graphite/SnO_2	0.693	18.47	0.52	6.71	31
Sample A	0.82	11.7	0.57	4.55	32
Sample B	0.89	12.1	0.60	5.73	
PANI	0.66	16.2	0.58	6.3	35
PEDOT-Ni_2P-3	0.75	13.91	0.68	7.14	36
PEDOT-Co_2P-3	0.75	13.61	0.68	6.85	
44 mM of $Pb(S_2COEt)_2$	7.32	14.283	0.48	4.71	38
66 mM of $Pb(S_2COEt)_2$	7.30	13.981	0.45	4.43	

Table XI Effect of temperature on the performance and interfacial resistance of the DSSC[42]

Temperature (°C)	V_{oc} (V)	J_{sc} (mA cm^{-2})	FF	PCE (%)	R (Ω)
25	0.632	8.212	0.53	2.581	282.40
35	0.627	7.432	0.55	2.567	251.32
45	0.534	7.003	0.56	2.085	40.340
55	0.475	6.745	0.54	1.659	35.061

Table XII Photovoltaic parameters of N719 dye-sensitized ZnO SCs at various sensitization temperatures[43]

Temperature (°C)	V_{oc} (V)	J_{sc} (mA cm^{-2})	FF	η (%)
5	0.74	1.81	0.38	0.51
25	0.75	2.72	0.69	1.16
40	0.87	4.29	0.61	2.26
60	0.78	8.79	0.32	2.19

The electrode's polarity influences the nature of the array. It was observed that $MoO_3$was adorned with rutile forms when the molybdenum electrode was polarized positively.[44] Temperature is a parameter that affects the performance of vertically aligned ZnO nanorod (NR)-based DSSC. With a current density of 2.06 mA cm^2, open-circuit voltage of 0.608 V, and fill factor of 0.46 at a temperature of 25 °C, the best photovoltaic response was observed with PCE of 0.57%, as mentioned in Table XIII.[45]

Table XIII Performance parameters of a DSSC at 100 mW cm^{-2} at distinct temperatures[45]

Temperature (°C)	V_{oc} (V)	J_{sc} (mA cm^{-2})	FF	η (%)
25	0.608	2.06	0.46	0.57
35	0.583	1.98	0.46	0.53
45	0.555	1.94	0.46	0.46
55	0.537	1.91	0.45	0.46
65	0.513	1.79	0.44	0.40
75	0.491	1.71	0.44	0.37

Challenges and Limitations

DSSCs are a simple, economical, and environmentally friendly method that may be applied for diverse applications. To enable DSSC fabrication for commercial use, their overall performance must be enhanced. The efficiency of the flow of electrons is affected by the coated film thickness. Although the doctor blade technique is mainly used for coating the conductive surface of counter electrodes, since it is a manual process, it may not always be optimal for coating. Therefore, alternative methods, including screen printing and spin coating, should be considered for the optimization and improved performance of counter electrodes. Carbonaceous materials are extremely durable and exhibit superior electrochemical characteristics, whereas conducting polymers exhibit excellent conductivity and low charge transfer resistance. Studies are needed to examine carbon-based

polymer composites for counter electrodes with various structural morphological features for fabricating extremely effective and affordable counter electrodes for efficient DSSCs. It has also been observed that the performance of the DSSCs declines with the increase in cell temperature. The decrease in cell efficiency can be attributed to the drop-off in the open-circuit voltage of the cell. The increased temperature also decreases the charge recombination resistance because of a decrease in electron lifetime.

Conclusion

Lab-scale experiments in solar technology are giving way to commercialization. New fabrication methods are being developed to increase productivity and lower costs. Due to their simple construction, low cost, and high efficiency, DSSCs have attracted the attention of researchers over the past 10 years and demonstrate enormous potential. The improvement of the counter electrode for DSSCs is discussed in this article, along with advancements in DSSCs. Researchers have synthesized and studied a number of materials as counter electrode in DSSCs, including metal oxides and metal sulfides, carbonaceous materials, and conducting polymers with diverse morphological features. These materials are addressed in the article.

Regarding photocatalytic activity, photo-conversion efficiency, low charge transfer resistance, and stability, certain materials demonstrated better photoelectric response and PCE, but were still not comparable to the Pt-based counter electrode. SWCNH/SSrGO and N-CCS have great catalytic performance in carbonaceous materials, according to their photovoltaic performance and electrical properties. Although PEDOT-Ni_2P-3 has demonstrated a strong photoelectric response when used as conducting polymer to reduce triiodide, it is costly as compared with platinum material. Various sulfides have demonstrated excellent catalytic activity because of their large surface area and lower charge transfer resistance. Carbonaceous materials and conducting polymers have shown effective performance with higher current density values. Consequently, various conducting polymer- and carbon-based materials must be studied with various morphology, synthesis methods, and fabrication procedures to increase photovoltaic performance. With further advances in the structural morphology of CE materials, it will be possible to improve the photo-conversion efficiency of photoactive materials for wider use as counter electrodes.

Acknowledgments Authors are highly thankful to Centre for Energy and Environment, Delhi Technological University for providing the basic facility for this review compiling.

Authors' Contribution GR Methodology, Writing—original draft and Investigation. AK Conceptualization, Writing—review and editing, Visualization and Supervision. AKS Investigation and editing. KNS Investigation and editing. BCM Conceptualization and Supervision.

Funding There is no funding.

Availability of Data and Materials Here critical analysis has been done with reference to earlier research work. This is a kind of comprehensive review. Hence, there are no data used.

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.

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

References

1. G. Richhariya, A. Kumar, P. Tekasakul, and B. Gupta, Natural Dyes for Dye Sensitized Solar Cell: A Review. *Renew. Sustain. Energy Rev.* 69, 705 (2017).
2. H. Nejadasad, A. Piri, and H. Zarei, The Effect of Carbon Nanotubes on the Efficiency of Dye Sensitized Solar Cells Based on TiO_2 Nanorods. *Optik* 142, 211 (2017).
3. M. Karpacheva, F.J. Malzner, C. Wobill, A. Buttner, E.C. Constable, and C.E. Housecroft, Cuprophilia: Dye Sensitized Solar Cells with Copper(I) Dyes and Copper(I)/(II) Redox Shuttles. *Dyes Pigm.* 156, 410 (2018).
4. Y. Jiao, L. Mao, S. Liu, T. Tan, D. Wang, D. Cao, B. Mi, Z. Gao, and W. Huang, Effects of Meta Or Para Connected Organic Dyes for Dye-Sensitized Solar Cell. *Dyes Pigm.* 158, 165 (2018).
5. G. Richhariya, and A. Kumar, Fabrication and Characterization of Mixed Dye: Natural and Synthetic Organic Dye. *Optic Mater.* 79, 296 (2018).
6. B. Tripathi, P. Yadav, and M. Kumar, Charge Transfer and Recombination Kinetics in Dye Sensitized Solar Cell Using Static and Dynamic Electrical Characterization Techniques. *Sol Energy* 108, 107 (2014).
7. M. Habib, N.N. Ghosh, R. Sarkar, A. Pramanik, P. Sarkar, and S. Pal, Controlling the Charge Transfer and Recombination Dynamics in Hollow ZnO QD Based Dye Sensitized Solar Cell: An Insight from Abinitio Simulation. *Chem. Phys. Lett.* 709, 21 (2018).
8. S. Sarker, H.W. Seo, F.O. Bakar, M.A. Aziz, and D.M. Kim, Facile and Rapid Preparation of Platinum Counter Electrodes for Dye-Sensitized Solar Cells. *J. Photochem. Photobiol. A Chem.* 321, 122 (2016).
9. S. Venkatesan, E.S. Darlim, I.P. Liu, and Y.L. Lee, Performance Enhancement Effects of Dispersed Graphene Oxide Sponge Nanofillers on the Liquid Electrolytes of Dye Sensitized Solar cells. *Carbon* 132, 71 (2018).
10. H. Jing, Y. Shi, W. Qiu, D. Wu, X. Song, Y. An, and C. Hao, Onion-Like Graphitic Carbon Covering Metallic Nanocrystals Derived from Brown Coal as A Stable and Efficient Counter Electrode for Dye-Sensitized Solar Cells. *J. Power Source.* 414, 495 (2019).
11. A. Ramar, R. Saraswathi, A.T.E. Vilian, S.M. Chen, and F.M. Wang, Poly Isothianaphthene/Graphene Nanocomposite as A New Counter Electrode Material for High Performance Dye Sensitized Solar Cell. *Synth. Metals.* 230, 58 (2017).
12. S.S. Nemala, P. Kartikay, S. Prathapani, H.L.M. Bohm, P. Bhargava, S. Bohm, and S. Mallick, Liquid Phase High Shear

Exfoliated Graphene Nanoplatelets as Counter Electrode Material for Dye-Sensitized Solar Cells. *J. Colloid Interface Sci.* 499, 9 (2017).
13. J.S. Choia, H.B. Parkb, O.J. Yoonc, and H.J. Kimb, Investigation on the Role of Graphene Oxide Sheet-Platinum Composite Counter Electrode in Dye Sensitized Solar Cell. *Thin Solid Films* 745, 139098 (2022).
14. G.N. Ngubeni, O. Akinbami, L. Mxakaza, S. Nkabinde, T. Kolokoto, F. Otieno, M.J. Moloto, K.P. Mubiayi, and N. Moloto, Evaluating the Effect of the Substrate on the Electrocatalytic Performance of Cu_2ZnSnS_4 and $Cu_2ZnSnSe_4$ Counter Electrodes in Dye-Sensitized Solar Cells. *Thin Solid Films* 745, 139099 (2022).
15. S.C. Yadav, A. Sharma, R.S. Devan, and P.M. Shirage, Role of Different Counter Electrodes on Performance of TiO_2 Based Dye-Sensitized Solar Cell (DSSC) Fabricated with Dye Extracted from Hibiscus Sabdariffa as Sensitizer. *Optic Mater.* 124, 112066 (2022).
16. P. Vijayakumar, M. Senthil Pandian, A. Pandikumar, and P. Ramasamy, A Facile One-Step Synthesis and Fabrication of Hexagonal Palladium-Carbon Nanocubes (H-Pd/C NCs) and Their Application As An Efficient Counter Electrode for Dye-Sensitized Solar Cell (DSSC). *Ceram. Int.* 43, 8466 (2017).
17. S.S. Nemala, P. Kartikay, S. Prathapani, H.L.M. Bohm, P. Bhargava, S. Bohm, and S. Mallick, Liquid Phase High Shear Exfoliated Graphene Nanoplatelets as Counter Electrode Material for Dye Sensitised Solar Cells. *J. Colloid Interface Sci.* 499, 9 (2017).
18. S.S. Nemala, K.S. Aneja, P. Bhargava, H.L.M. Bohm, S. Mallick, and S. Bohm, Novel High Pressure Airless Spray Exfoliation Method for Graphene Nanoplatelets As a Stable Counter Electrode in DSSC. *Electrochim. Acta.* 285, 86 (2018).
19. S.S. Nemalaa, P. Kartikaya, R.K. Agrawala, P. Bhargava, S. Mallicka, and S. Bohm, Few Layers Graphene Based Conductive Composite Inks for Pt Free Stainless Steel Counter Electrodes for DSSC. *Sol. Energy* 169, 67 (2018).
20. W. Wang, J. Yao, X. Zuo, and G. Li, High Efficiency Nitrogen Doped Core Shell Carbon Spheres as Counter Electrodes for Dye-Sensitized Solar Cells. *Mater. Lett.* 227, 172 (2018).
21. A.B. Suriani, A. Muqoyyanah, M.H. Mohamed, N. Mamat, I.M. Hashim, M.F. Isa, M.I. Malek, A.R. Kairi, and M.K.A. Mohamed, Improving the Photovoltaic Performance of DSSCs Using a Combination of Mixed Phase TiO_2 Nanostructure Photoanode and Agglomerated Free Reduced Graphene Oxide Counter Electrode Assisted with Hyperbranched Surfactant. *Optik* 158, 522 (2018).
22. M. Gurulakshmi, A. Meenakshamma, K. Susmitha, N. Charanadhar, V.V.S.S. Srikanth, S. Narendra Babu, Y.P. Venkata Subbaiah, M. Katta Venkateswarlu, and Raghavender, A Transparent and Pt-Free All-Carbon Nanocomposite Counter Electrode Catalyst for Efficient Dye Sensitized Solar Cells. *Sol. Energy* 193, 568 (2019).
23. M. Younas, M.A. Gondal, M.A. Dastageer, and K. Harrabi, Efficient and Cost Effective Dye-Sensitized Solar Cells Using MWCNT-TiO_2 Nanocomposite as Photoanode and MWCNT as Pt-Free Counter Electrode. *Sol. Energy* 188, 1178 (2019).
24. H. Wang, J. Gao, J. Zhu, J.Y. Ma, H. Zhou, J. Xiao, and M. Wu, Design Bifunctional Nitrogen Doped Flexible Carbon Sphere Electrode for Dye-Sensitized Solar Cell and Supercapacitor. *Electrochim. Acta.* 334, 135 (2020).
25. S. Tontapha, W. Sang, T. Promgool, S. Kanokmedhakul, W. Maiaugree, E. Swatsitang, V. Homrahad, and V. Amornkitbumrung, Electrocatalytic Activity of Disulfide/Thiolate with Graphene Nanosheets As An Efficient Counter Electrode for DSSCs: A DFT Study. *Mater. Today Commun.* 22, 100742 (2020).
26. R.S. Kumar, M. Balu, S. Ramakrishnan, P.C. Ramamurthy, S.K. Batabyal, D. Kumaresan, and N.K. Kothurkar, Molybdenum Disulfide/Reduced Graphene Oxide Hybrids with Enhanced Electrocatalytic Activity: An Efficient Counter Electrode for Dye-Sensitized Solar Cells. *J. Electroanal. Chem.* 847, 113236 (2019).
27. H.J. Ahn, J.S. Lee, H.S. Kim, I.T. Hwang, J.H. Hong, J. Shin, and C.H. Jung, Fabrication of Large Pt Nanoparticles-Decorated rGO Counter Electrode for Highly Efficient DSSC. *J. Ind. Eng. Chem.* 65, 318 (2018).
28. M. Younas, M.A. Gondal, M.A. Dastageer, and U. Baig, Fabrication of Cost Effective and Efficient Dye Sensitized Solar Cells with WO_3-TiO_2 Nanocomposites as Photoanode and MWCNT as Pt-Free Counter Electrode. *Ceram. Int.* 45, 936 (2019).
29. A. Kamppinen, K. Aitola, A. Poskela, K. Miettunen, and P.D. Lund, Stability of Cobalt Complex Based Dye Solar Cells with PEDOT and Pt Catalysts and Different Electrolyte Concentrations. *Electrochim. Acta.* 335, 135652 (2020).
30. M.A. Ehsan, M.Y. Ounas, A. Rehman, M. Altaf, M.Y. Khan, A. Al-Ahmed, S. Ahmad, and A.A. Isabb, Synthesis and Utilization of Platinum(II) Dialkyldithiocarbamate Precursors in Aerosol Assisted Chemical Vapor Deposition of Platinum Thin films as Counter Electrodes for Dye-Sensitized Solar Cells. *Polyhedron* 166, 186 (2019).
31. M.A.K.L. Dissanayake, J.M.K.W. Kumari, G.K.R. Senadeera, T. Jaseetharana, J. Weerasinghe, H. Anwar, and A. Low-Cost, Vein Graphite/Tin Oxide Nanoparticles Based Composite Counter Electrode for Efficient Dye-Sensitized Solar Cells. *Mater. Sci. Eng. B* 273, 115440 (2021).
32. P. Mahajan, R. Datt, V. Gupta, and S. Arya, Synthesis and Characterization of ZnO@WO_3 Core/Shell Nanoparticles as Counter Electrode for Dye Sensitized Solar Cell. *Surf. Interfac* 30, 101920 (2022).
33. Q. Han, Z. Hu, H. Wang, Y. Sun, J. Zhang, L. Gao, and M. Wu, High Performance Metal Sulfide Counter Electrodes for Organic Sulfide Redox Couple in Dye-Sensitized Solar Cells. *Mater. Today Energy* 8, 1 (2018).
34. S.S. Nemala, K. Mokurala, P. Bhargava, and S. Mallick, Titania Nanobelts as a Scattering Layer with Cu_2ZnSnS_4 as a Counter Electrode for DSSC with Improved Efficiency. *Mater. Today Proc.* 5, 23351 (2018).
35. M.O. Karakus, M.E. Yakisiklier, A. Delibas, E. Ayyildiz, and H. Cetin, Anionic and Cationic Polymer Based Quasi Solid State Dye Sensitized Solar Cell with Poly(Aniline) Counter Electrode. *Sol. Energy* 195, 565 (2020).
36. Y. Di, S. Jia, N. Li, C. Hao, H.Z.S. Hu, and H. Liu, Electrocatalytic Films of PEDOT Incorporating Transition Metal Phosphides as Efficient Counter Electrodes for Dye Sensitized Solar Cells. *Sol. Energy* 189, 8 (2019).
37. M.U. Shahid, N.M. Mohamed, A.S. Muhsan, R. Bashiri, A.E. Shamsudin, and S.N.A. Zaine, Few-Layer Grapheme Supported Polyaniline (PANI) Film as a Transparent Counter Electrode for Dye Sensitized Solar Cells. *Diam Relat. Mater.* 94, 242 (2019).
38. U. Mehmood, and A.U.H. Khan, Spray Coated PbS Nano-Crystals as An Effective Counter-Electrode Material for Platinum Free Dye Sensitized Solar Cells (DSSCs). *Sol. Energy* 193, 1 (2019).
39. S. Yun, X. Zhou, Y. Zhang, C. Wang, and Y. Hou, Tantalum Based Bimetallic Oxides Deposited on Spherical Carbon of Biological Origin for Use as Counter Electrodes in Dye Sensitized Solar Cells. *Electrochim. Acta* 309, 371 (2019).
40. G. Syrrokostas, K. Govatsi, G. Leftheriotis, and S.N. Yannopoulos, Platinum Decorated Zinc Oxide Nanowires As An Efficient Counter Electrode for Dye Sensitized Solar Cells. *J. Electroanal. Chem.* 835, 86 (2019).
41. K. Subalakshmi, K. Ashok Kumar, O. Prayer Paul, S. Saraswathy, A. Pandurangan, and J. Senthilselvan, Platinum free Metal Sulfide Counter Electrodes for DSSC Applications: Structural, Electrochemical and Power Conversion Efficiency Analyses. *Sol. Energy* 193, 507 (2019).

42. U. Mehmood, M.I. Malik, A.U.H. Khan, I.A. Hussein, K. Harrabi, and A.A. Ahmed, Effect of Outdoor Temperature on the Power-Conversion Efficiency of Newly Synthesized Organic Photosensitizer Based Dye-Sensitized Solar Cells. *Mater. Lett.* 220, 222 (2018).
43. A. Roy, S. Sundaram, and T.K. Mallick, Effect of Dye Sensitization's Temperature on ZnO Based Solar Cells. *Chem. Phys. Lett.* 776, 138688 (2021).
44. A. Khlyustova, N. Sirotkin, V. Titov, and A. Agafonov, Effect of Low-Temperature Underwater Plasma Produced of New Properties of Mo–Ti Mixed Oxide Composites for Electron Transport Layer in the Dye-Sensitized Solar Cells. *J. Alloys Compd.* 858, 157664 (2021).
45. I.G. Valls, Y. Yu, B. Ballesteros, J. Oro, and M.L. Cantu, Synthesis Conditions, Light Intensity and Temperature Effect on the Performance of ZnO Nanorods-Based Dye Sensitized Solar Cells. *J. Power Source* 196, 6609 (2011).

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

PAPER • **OPEN ACCESS**

Emission performance assessment of in-use diesel-driven cars

To cite this article: Abhinav Pandey *et al* 2022 *IOP Conf. Ser.: Earth Environ. Sci.* **1084** 012041

View the article online for updates and enhancements.

You may also like

- Study on deformation of abutment and the influence on high arch dam during impoundment
X W Wang, J R Xu, L J Xue et al.
- Design of telescopic stub tuner of 1 5/8 transmission line
Atul Varia, Raj Singh, S V Kulkarni et al.
- A Comparison of Four Algorithms for Land-Use Classification Based on Landsat 8 OLI Image
Yanghua Zhang, Hu Zhao and Jianbo Ni



This content was downloaded from IP address 14.139.251.98 on 25/10/2022 at 05:20

Emission performance assessment of in-use diesel-driven cars

Abhinav Pandey[1], Rajeev Kumar Mishra[2], Govind Pandey[3]

[1]Department of Environmental Engineering, Delhi Technological University, India
[2]Department of Environmental Engineering, Delhi Technological University, India
[3]Civil Engineering Department, Madan Mohan Malaviya University of Technology, India)

Email: navabhigkp@gmail.com

Abstract. The overall fleet of light-duty vehicles in most of the developing countries is characterized by various manufacturers of diesel-driven passenger cars. There is a need of a thorough assessment of the actual conformance levels of such cars towards in-use emission norms, particularly the diesel-driven ones, in consideration of their high proportion in the overall privately-owned passenger cars fleet. In this context, the study conducted on 400 number of light-duty diesel cars reported that the compliance levels for BS (Bharat Stage) III diesel cars are better than that for BS IV, although not varying significantly (90 % in case of former vs. 80 %). Based on the findings, there is a need to include vehicle-related parameters during emission certification, such as, age, mileage and maintenance records along with emission measurements while deciding upon phasing-out of vehicles rather only based upon age (10 years of allowable registration life for diesel cars in the state of Delhi). Based on the results, the inclusion of mass emission-based testing in the overall transport emission assessment and reduction policy of the developing countries is highly recommended.

Keywords: Emission performance, In-use emission norms, Inspection/Maintenance program, Maintenance category, Tailpipe emission, Environmental policy.

1. Introduction

The exponential and advertent surge in the motor vehicle density (number of vehicles per thousand of human population) in most of the urban settlements worldwide looks to have surpassed the expansion rate of the cities itself. Such expansion is more manifest in the urban areas of the developing countries which are often marked by the relatively faster growth in few cities compared to the remaining parts. Although, there has been an apparent decline in the sales of diesel cars considering the preceding decade, a substantial chunk of the 'in-use vehicles' is still characterized by such cars (Pandey et al. 2022a). Most of the times, this high proportion of diesel cars is aggregated in the megacities (including suburbs) in contrast to several other urban areas letting such megacities emit a larger fraction of air pollutants at global scale (Lawrence et al. 2007; Butler 2013; Pandey et al. 2022a). The worldwide passenger car (PC) sales have witnessed a moderate-to-very high growth of 21.16 % (in 2019 compared to 2010) and 17.97 % (in 2019 against that in 2005) while African nations recorded the lowest sales volume. Compared to a total of 45.25 million passenger cars sold globally in 2005, 2019 witnessed the numbers as 90.42 showing almost doubling of numbers in only 14 years' span (OICA 2019; Pandey et al. 2022b). The infrastructure development, unfortunately, has not been able to equalize with this exponentially

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

rising degree of motorization, aggravating the problem of poor urban air quality (Wang et al. 2019; Jaikumar et al. 2017), unplanned growth and changing meteorology (Bai et al. 2017).

Until recently, driven by their advantages like better thermal efficacy, higher fuel mileage, and relatively higher notch of consistency, higher magnitude of torque compared to other engines, the diesel engine remained a prime choice. On the other hand, as they also, tend to emit harmful exhaust emissions, such as CO, unburned hydrocarbons HC, NO_X, SO_2 and smoke or PM, more or less similar to the other extant engine type. Further, as the NO_X and SE (smoke) are relatively higher in tailpipe exhaust from the diesel engines compared to petrol engines, their environmental performance has been subject to strict policy actions as well (Karthikeyan and Prathima 2016; Sharma and Marechal, 2020; Pandey et al. 2022a).

According to the global health report, over 4.2 million deaths annually in developing countries alone, *inter-alia* several cases of pulmonary ailments are being attributable to the deteriorating urban ambient air quality (WHO 2019). Moreover, both long-term and short-term (chronic and acute in other terms) exposures to vehicular exhaust emission (VEE) has been triggering various epidemiological consequences, ranging from cardiovascular disease, waning pulmonary function, oncological cases and even mortality in extreme instances (Ogunseye et al. 2018; Rice et al. 2015; Khreis et al. 2016). Various researches in the past have evidenced vehicular emission (VE) to be contributing to approximately 70–80% of air pollution in mega cities of developing nations. Such high apportionment is being credited not only to the increasing sales of passenger cars, but also to the presence of significantly large number of older cars coupled with poor maintenance history, cars having clocked higher mileage, inadequate access and circulation infrastructure and inferior fuel quality (Singh et al. 2007; Pandey et al. 2016; Pandey el al. 2022a, b). Among the criteria (air) pollutants, CO is the major pollutant emanating from the transport sector and contributing to approximately 90% of overall emissions with Hydrocarbons following closely. Further, the vehicular emissions are known to contribute to the small air particles such as black carbon (BC), nitrogen oxides (NOx), and particulate matter (PM_{10}, $PM_{2.5}$ and PM_1) as prominent air pollutants (Rao et al. 2018).

As part of overall environmental policy, the I/M (Inspection and Maintenance) program in India mandates all the in-use diesel-driven passenger cars to acquire and possess the Pollution Under Control (PUC) certificate at a fixed periodicity. Such certificate is issued in the name of the owner/registered car if it is found to comply to the relevant smoke (exhaust) emission norm conducted using the smoke testing under free acceleration mode (FAST). These emission checks must be carried out only at the government-authorized centres (PUC centres) which use approved smoke meters (CPCB 2010). Previous relevant studies have pointed towards the need to explore and characterize the emission quality of vehicles with a fairly larger and heterogeneous dataset with more model years, and in the context of more parameters concerning vehicles, such as, age, mileage clocked, maintenance history and records, in-use emission norms and various engine-related traits as the key ones (Pandey et al. 2016).

Governments, policymakers, and managers have been affecting and trying to strengthen various policy measures to improve urban air quality by maintaining ambient air pollution levels below the specified national and international standards. At-the-source control of vehicular pollution is among the most promising practices for pollution abatement, but this approach requires a very stringent administration system, policy revisits and re-strengthening, public awareness and support, and monitoring for implementation (Gulia et al. 2020).

Except US, EU and other developed countries plus China and Brazil, the emission certification policies for in-use passenger cars in most developing countries are struggling to meet the overall vehicular emission reduction goals. There is lack of a comprehensive exploratory analysis of exhaust emission compliance of the diesel-driven light-duty passenger cars to the in-use emission norms required covering a large and heterogeneous make-wise dataset. Therefore, in the present study, 400 numbers of private registration type diesel cars appearing at PUC (Pollution Under Control) centres for emission recertification, were assessed for their tailpipe smoke emission (SE) in HSU (Hartridge Scale Unit) terms in a collaborated testing program. The testing criteria was in line with CVMR (Central Motor Vehicle Rules) prescribed by the Ministry of Road Transport and Highways (MoRTH), Government of

India. The protocol, at present, requires the diesel vehicles to be tested under stationary and no-load conditions and the PUC certificate is issued in case of conformance to the in-use emission limits.

2. Materials and methods

2.1. Vehicular emission testing

The vehicle testing schedule covered 400 numbers of diesel-driven light-duty passenger cars. The cars appearing for emission compliance certification at respective PUC (Pollution-Under-Control) centres in the NCT (National Capital Territory) of Delhi, India were checked for smoke emission from the tailpipe in FAST mode. The compliance level of the diesel cars was checked against respective in-use norms applicable in India **(Table 1)**. The analysis was carried out for three dataset scenarios – whole, top 5 makes and top 13 models.

Table 1. Sample characteristics at a glance

Emission norm	HSU @ FAST
BS (Bharat Stage) IV	50
BS (Bharat Stage) III	65

2.2. Instrumental set-up and data analysis tool

A smoke meter of the Ozone make (model no. Oz-Gas-02) was deployed to measure smoke emitted by the diesel vehicles (in HSU terms) under the FAST mode from the tailpipe of the cars. Further, the SPSS package (version 23) was used in the present study for the emission data compilation, analysis and processing along with the MS-Excel program as a support (input) tool.

3. Results and discussion

The status of compliance of diesel cars for different dataset scenarios is presented in Figure **1**, followed by **Fig. 2** and **Fig. 3** depicting the same for make and model-wise status, respectively.

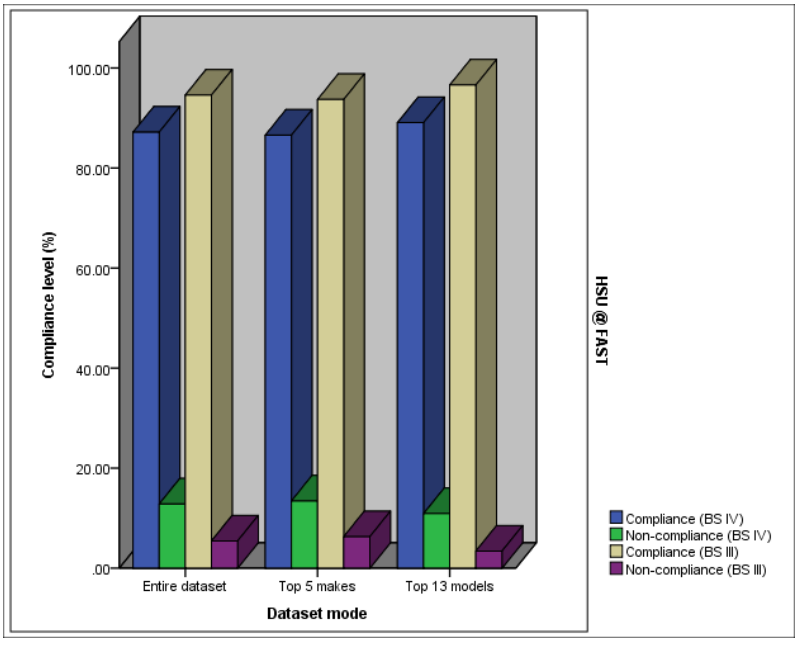


Figure 1. Emission compliance levels for diesel-driven cars (all datasets)

As evidenced, all the datasets reveal a very high level of compliance to both BS IV and BS III norms for HSU in the FAST mode of testing. As the BS III benchmark is higher than that of BS IV (65 HSU against 50 HSU), compliance of diesel cars towards BS III is significantly and logically so, higher in all dataset conditions.

In terms of makes, the MML has lowest degree of compliance amongst the 5 makes closely followed by TKMPL whereas HMIL depicts the highest degree of compliance (over 90%), Figure 2.

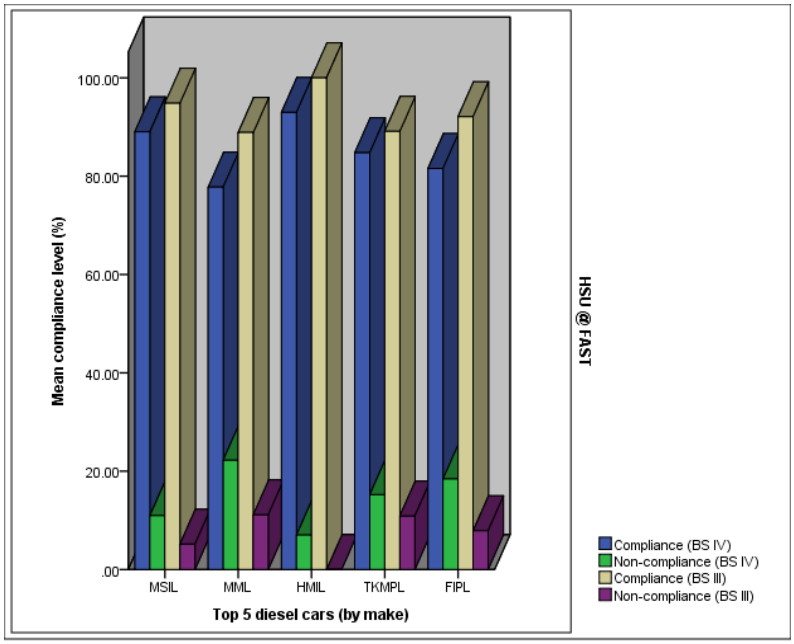


Figure 2. Emission compliance levels for diesel-driven cars (top 5 makes)

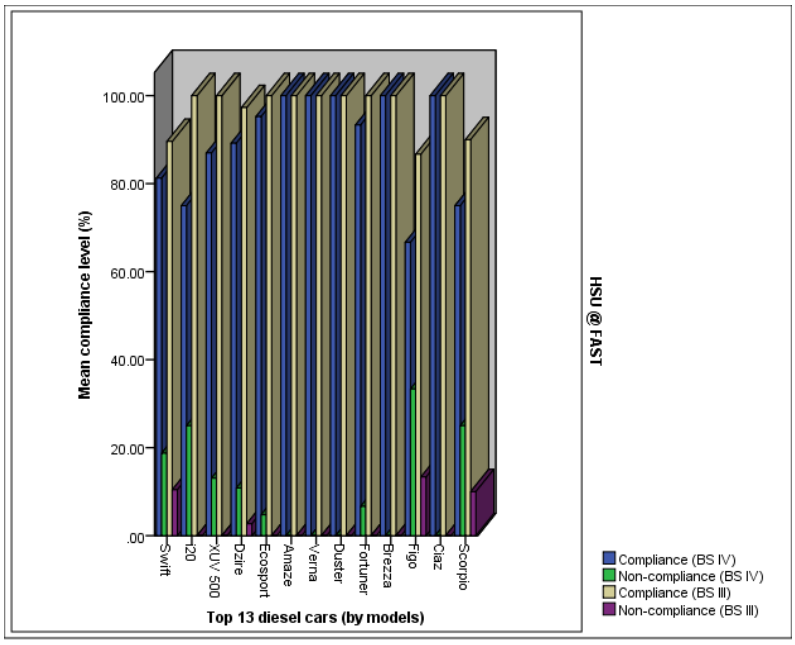


Figure 3. Emission compliance levels for diesel-driven cars (top 13 models)

The top 13 models' dataset scenario fares better than other datasets in terms of non-compliance to both BS III and BS IV norms returning the lowest values. Amaze, Verna, Duster, Brezza and Ciaz depict the highest degree of compliances for both BS IV and BS II emission norms of HSU under FAST showing almost 100% compliance. Figo leads the list of lowest levels of compliance followed by Scorpio, i20 and Swift models (Figure 3).

4. Conclusion

The present study is an attempt to investigate the compliance levels of smoke emission from 400 diesel-driven light-duty passenger cars towards in in-use emission norm. The cars tested were registered in the different regional transport offices (RTOs) across the state of Delhi, India. The study reported that the diesel cars reveal a very high level of compliance to both BS IV and BS III norms for HSU in FAST mode of testing. While make-wise, HMIL performs the best towards both the emission standards, followed by MSIL and TKMPL, MML and FIPL have higher percentage of non-compliance, especially for BS IV. Amaze, Verna, Duster, Brezza and Ciaz depict the highest degree of compliances for both BS IV and BS II emission norms of HSU under FAST showing almost 100% compliance. In model terms, Figo leads the list of lowest levels of compliance followed by Scorpio, i20 and Swift models. Although such compliance levels seem to be encouraging for the car manufacturers in justification of the fact that the improved vehicle and engine technologies have contributed to the less-emitting engines, the increasing ambient levels of both gaseous and particulate pollutants in Delhi and in other major economies worldwide present a contrasting picture. This issue or the gap needs to be dealt by the future studies in the concurrent area. Nevertheless, the policymakers should take into account the more realistic real-time driving tests to correlate the results reported using the in-situ laboratory-based emission tests. It is emphasized that in view of the introduction of and insistence over the stricter emission standards worldwide (mostly for the new vehicles), the future research works need to be carried out with vehicle model or fuel-based emission datasets. Such research works also need to keep in mind the real-world driving conditions and emission generation thereof.

MSIL – Maruti Suzuki India Limited; MML – Mahindra Motors Limited; HMIL – Hyundai Motors India Limited; TKMPL – Toyota Kirloskar Motors Private Limited; FIPL – Ford India Private Limited; NOx – Oxides of Nitrogen; SO_2 – Sulphur Dioxide; SE – Smoke Emission; PM – Particulate Matter; CPCB – Central Pollution Control Board; EU – European Union; US – United States (of America); HSU – Hartridge Smoke Unit; I/M – Inspection & Maintenance; IBM – International Business Machines; SPSS – Statistical Package for Social Sciences; VEE – Vehicular Exhaust Emission; VE – Vehicular Emission; FAST – Free Smoke Acceleration Testing

References

[1] A. Pandey, G. Pandey, and R.K. Mishra, An in situ exploratory analysis of diesel cars' emission: way forward on policy evaluation, Environmental Science and Pollution Research, 2022.
[2] M.G. Lawrence, T.M. Butler, J. Steinkamp, B.R. Gurjar, and J. Lelieveld, Regional pollution potentials of megacities and other major population centres, Atmospheric Chemistry and Physics, 7, 2007, 3969–3987.
[3] T. Butler, Urban and global effects of megacity pollution. In: Barnes I., Rudziński K. (eds) Disposal of dangerous chemicals in urban areas and mega cities, 2013 NATO Science for Peace and Security Series C: Environmental Security. Springer, Dordrecht.
[4] OICA (International Organization of Motor Vehicle Manufacturers) 2019, *Global sales statistics of passenger cars,* https://www.oica.net/category/sales-statistics/ Accessed 10 January 2020.
[5] A. Pandey, R.K. Mishra; and G. Pandey, Investigating exhaust emissions from in-use passenger cars: exploratory analysis and policy outlook, Journal of Environmental Engineering (American Society of Civil Engineering), 2022, 148(7), 04022035.

[6] J. Wang, Q. Wu, J. Liu, H. Yang, M. Yin, S. Chen, P. Guo, J. Ren, X. Luo, W. Linghu, and Q. Huang, Vehicle emission and atmospheric pollution in China: problems, progress, and prospects. PeerJ, 16,7, 2019, e6932 PMID: 31143547; PMCID: PMC6526014
[7] R. Jaikumar, S.M. Shiva Nagendra, and R. Sivanandan, Modal analysis of real-time, real world vehicular exhaust emissions under heterogeneous traffic conditions. Transportaiton Research Part D Transport and Environment, 54, 2017, 397–409
[8] X. Bai, T. McPhearson, H. Cleugh, H. Nagendra, X. Tong, T. Zhu, and Y.G. Zhu, Linking urbanization and the environment: conceptual and empirical advances. Annual Review of Environment and Resources, 42, 2017, 215–240.
[9] S. Sharma, and F. Marechal F, Carbon dioxide capture from internal combustion engine exhaust using temperature swing adsorption. Frontier in Energy Research 7, 2019, 143.
[10] S. Karthikeyan, and A. Prathima, Environmental effect of CeO_2 nano-additive on biodiesel, Energy Sources A: Recovery, Utilization and Environmental Effects, 38(24), 2016, 3673–3679.
[11] WHO (World Health Organization) 2019, *The global health observatory*, https://www.who.int/data/gho/data/themes/theme-details/GHO/air-pollution Accessed 7 January 2020.
[12] O.O. Ogunseye, G.R.E.E. Ana, D.C. Uhiara, and D.G. Shendell, Carboxyhaemoglobin levels among traders exposed to vehicular emissions in three motor parks in Ibadan, Nigeria, Journal of Environmental and Public Health, 2018, 1–8.
[13] M.B. Rice, P.L. Ljungman, E.H. Wilker, K.S. Dorans, D.R. Gold, J. Schwartz, P. Koutrakis, G.R. Washko, G.T. O'Connor, and M.A. Mittleman, Long-term exposure to traffic emissions and fine particulate matter and lung function decline in the Framingham Heart Study, American Journal of Respiratory and Critical Care Medicine, 191, 2015, 656–664.
[14] H. Khreis, K.M. Warsow, E. Verlinghieri, A. Guzman, L. Pellecuer, A. Ferreira, I. Jones, E. Heinen, D. Rojas-Rueda, N. Mueller, P. Schepers, K. Lucas, and M. Nieuwenhuijsen, The health impacts of traffic-related exposures in urban areas: understanding real effects, underlying driving forces and co-producing future directions, Journal of Transport and Health, 3(3), 2016, 249–267
[15] A.K. Singh, H.K. Gupta, K. Gupta, P. Singh, V.B. Gupta, and R.C. Sharma, A comparative study of air pollution in Indian cities, Bulletin of Environmental Contamination and Toxicology, 78, 2007, 411–416.
[16] A. Pandey, G. Pandey, and R.K. Mishra, Tailpipe emission from petrol driven passenger cars, Transportation Research Part D Transport and Environment, 44, 2016, 14–29.
[17] X. Rao, J. Zhong, R.D. Brook, and S. Rajagopalan, Effect of particulate matter air pollution on cardiovascular oxidative stress pathways, Antioxidants & Redox Signaling, 28(9), 2018, 797–818.
[18] CPCB (Central Pollution Control Board), 2010, Status of the vehicular pollution control programme in India. Central Pollution Control Board, Ministry of Environment, Forests and Climate Change, Delhi.
[19] S. Gulia, S.S. Nagendra, J. Barnes, and M. Khare, Urban local air quality management framework for non-attainment areas in Indian cities, Science of the Total Environment, 619, 2018, 1308–1318.
[20] S. Caserini, C. Pastorello, P. Gaifami, and L. Ntziachristos, Impact of the dropping activity with vehicle age on air pollutant emissions, Atmospheric Pollution Research, 4, 2013, 282–289.

Acknowledgements
The authors would like to express their thanks to the PUC centres in Delhi and the Advance Air and Acoustics Research Lab, Delhi Technological University (DTU) for their support in the emission measurement program, data collection and processing thereafter.

Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Estimation of Daily Groundwater Table Using Backpropagation Neural Network Model by Assessing Training Algorithm

[1]Varna Vishakar V, [2] Ayush Jain, [3]Zohaib Ahmed Khan,[4]Ayush Kumar

[1][2][3][4] Assistant Professor, Department of Civil Engineering, KIET Group of Institutions, Delhi-NCR, Meerut Road(NH-58), Ghaziabad.

[1]varnavishakar@gmail.com,[2]ayushjain2904@gmail.com,
[3]zohaib.khan@kiet.edu,[4]ayush.kumar@kiet.edu

Article Info
Page Number: 374 – 388
Publication Issue:
Vol. 71 No. 3s (2022)

Abstract

The Onça stream is a Brazilian waterway, a canal on the left bank of the Rio das Velhas and a tributary of the São Francisco River. Although it plays an important role in the management of groundwater levels in the Ribeirao Preto region, the final groundwater capacity is declining in many parts of Brazil in response to abstraction. Predicting and forecasting stream water level using simple but effective methods can provide a reliable tool for future management of groundwater. In the present study, the daily water level of a particular well from the year 2004 to 2014 was considered as the dataset. The adopted database was segregated as 70% for training the above-mentioned models and the remaining 30% data was utilized for model testing. One of the soft computing methods, Backpropagation Neural Network was utilized for the prediction of daily groundwater table. In this, BPNN model, various training algorithm were utilized and compared based on the value of coefficient of correlation (R)and the time taken. The model also assessed by various statistical parameters like root means square error (RMSE), mean total error (MAE), determination coefficient (R^2), Normalized mean biased error (NMBE). The results depicted that the developed model have good forecasting ability.

Keywords: Backpropagation neural network, Estimating, Groundwater table, Soft computing, Training algorithm.

Article History
Article Received: 22 April 2022
Revised: 10 May 2022
Accepted: 15 June 2022
Publication: 19 July 2022

1. Introduction

The weather condition variations, water pollution are the certain factors which elevates the water requirement across the globe. One of the most important sources of freshwater is the groundwater, however, the abstraction and climate changes diminish the groundwater

level**[3][21**]. Groundwater research is extremely important because surface water is scarce in arid and semi-arid countries. These naturally filtered liquids are sometimes mistaken for good drinking and flowing water. The importance of multivariate modelling occurred in the planning, construction and operation of Meteorological and hydrological systems are frequently used in water resource systems.(for example, precipitation, flow, and temperature) **[19][17][12]**. Artificial Intelligence approaches including Genetic Programming (GP), Adaptive Neuro Fuzzy Inference System (ANFIS), Artificial Neural Networks (ANN), and M5tree (M5T) have been understood in recent years: that it was used ANNs to estimate monthly groundwater level variations**[22][20].**ANN was used to estimate subsurface water levels; A hybrid ANN-numerical model for groundwater problems**[1][5].**ANN model was used to solve a challenging problem, In New Jersey, there is a real-world subsurface water management challenge**.[2][4][6**].ANN to estimate groundwater levels in China's West Jilin Province. ANN was used to investigate the effects of human activities on regional groundwater levels. ANNs were developed by et al. (2005) to predict potentiometric surface elevations accurately. **[19][11][9]**The primitive dispute in the water resource management is the meticulous prediction of groundwater level **[8].** As urban areas extend across the world, governments and organizers search for approaches to precisely anticipate the groundwater level so they can ensure the water request is met as the populace development continues to increment and furthermore as a vital factor when planning new urban areas**.[13][15].** The groundwater table has been estimated by using various modelling techniques [Artificial Neural Network is one of the supervised learning inspired from human brain's architecture. During the early 1970's the Backpropagation neural network was established but the fame of the BPNN was escalated by Rumelhart (1986). The shortage of water assets has driven cultivators to improve their water system procedures to give crops their accurate water prerequisites. The external climatic conditions were utilized in order to predict the leaf area index for a greenhouse transpiration model by adopting the BPNN model **[14][23].** The spatial evolution of sea ice of the Liaodong Bay was predicted with précised accuracy by the BP neural network model**[10]**. Ground source heat pump (GHSP) systems have been broadly utilized in private and business structures all through the world because of their elite and ecological kind disposition. The prediction of GHSP was carried out by adopting the BP model and the output was promising with higher accuracy**[18]**. The BP model was utilized as an evolutionary soft computing model for numerous real time problem **[7][16]**.In this study, the goal is to evaluate the different training algorithm in BP model in order to predict the daily groundwater table in the Rio das Velhas region in Brazil. The evaluation of this model was carried out by different statistical indices such as Root mean square error (RMSE), Mean Absolute error (MAE), Coefficient of determination (R^2), Normalized mean bias error (NMBE), Performance Index (PI), Nash–Sutclie coefficient (NS). In order to construct the BPNN model, MATLAB was utilized.

2. Study area & Data assemblage

The Ribeirão das Posses watershed is located in the south of Minas Gerais State, Brazil, and extends between geographic coordinates from 22°49'46" to 22°53'21" south and from 46°13'24" to 46°15'08" west. The study area is depicted in the following figure 1.The daily

groundwater table for a particular well in the above-mentioned location was observed from the year 2004 till 2014.

In order to model the nonlinear hydrological process, the selection of possible input variable was one of the disputes. In this study, the various input fusions were adopted to resolve this hydrological modelling issue. The observed data from the year 2004-2014 was calibrated. This model, utilized 70% of the data for training and 30% of the data for the testing, as there is no universal rule for the segregation of data (Taheri et al., 2019; Pham et al., 2019; Chen et al., 2019; Khosravi et al., 2019; Khozani et al., 2019). The data was normalized between 0 and 1 by using the following equation (1)

$$\text{normalized value} = \frac{\text{value}-\text{min.value}}{\text{max.value}-\text{min.value}} \quad (1)$$

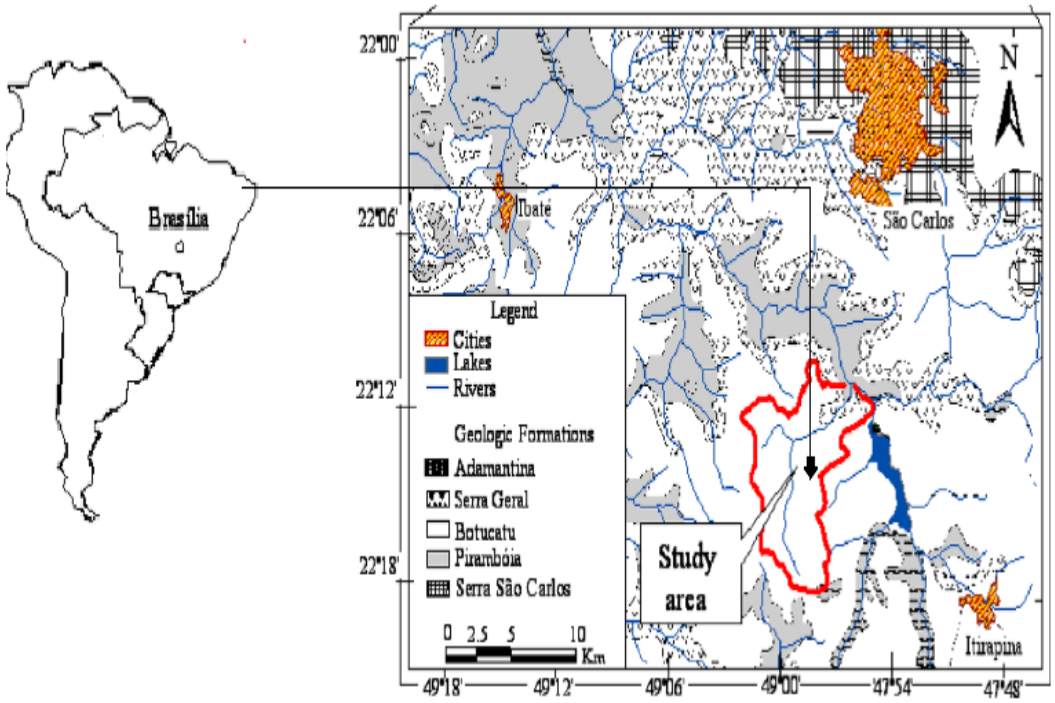


Figure 1: Location of the study area.

3. Methodology

Artificial neural networks (ANNs) are one of the computing techniques and systems that can derive new information by learning from the properties of the human brain's ability to create and discover new information. They were developed with the help of the human brain's ability to create and discover new information.The goal is to be able to execute without assistance. Inspired by artificial neural networks, the human brain is the consequence of learning process mathematical modelling.The feed-forward-backpropagation ANN methodology, which works on the idea of backpropagation, is the most extensively used ANN method. An ANN model has three layers: input, hidden, and output.

Standard patterns are required in ANN in order to build an essential relationship between the data pair of inputs and outputs. Additionally, the predictability of a skilled ANN should also be examined for the situations which were not the part of the neural network training.

The ANN models generally used to build up prescient models in light of its self-learning, nonlinearity, and self-assertive capacity estimation capacity. The BPNN model is one of the subsets of the ANN, which has typically input, hidden and the output layer. The main pro of this BP model was the reverse transmission of signals and errors. The typical architecture of the BPNN model was depicted in the following figure 2.

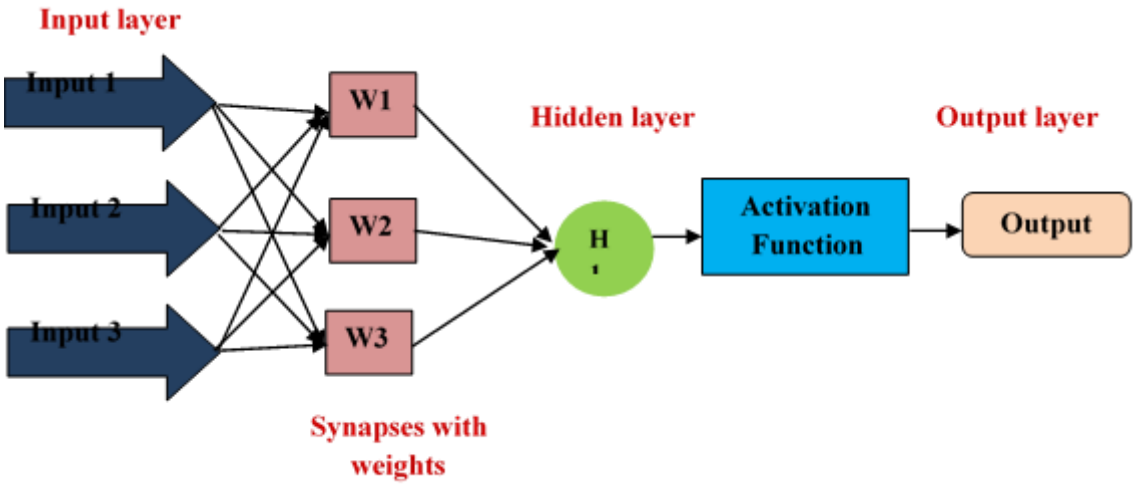


Figure 2: Typical architecture of BPNN model.

In the back propagation model, the preliminary weights were set arbitrarily and later it was altered as per the disparities between the training and the output data. The standard BP neural network utilizes an algorithm of descent gradient, and the network weights are changed contrarily along the gradient of the function. Consequently, the anticipated values continue to draw nearer to the observed value (Feng et al., 2020). In the initial stage, the BPNN model, promulgate the inputs forward and the yield of the various layers can be computed by using the following equation (2) (Jana et al., 2018).

$$G^{n+1} = f^{n+1}(w^{n+1}.G^{n+1} + b^{n+1}) \quad (2)$$

where G- is the output, f is the transfer function, w is referred as the weight matrix for various layers; n=0, 1…,M-1, M is the number of layers; and b is termed as the bias matrix for the considered layers.,The next stage in BP is to disseminate the sensitivities backward through the network initiating from the final layer.

$$s^M = 2K^M(x)^M(t - S) \quad (3)$$

where, s is the sensitivity, x is the net input and t is considered to be the target

$$K^M = [f^n(x_1^n) \ \dots \ 0 \ \vdots \ f^n(x_2^n) \ \vdots \ 0 \ 0 \ f^M(x_{s^n}^n) \quad]$$

Then the last layer of the sensitivities is then back propagated to the initial layer by the and it can be expressed as per the following equation (4)

$$s^n = F^n(x^n)(W^{n+1})^T . s^{n+1} \quad (4)$$

The biases and the weights will be upgraded in the concluding step by adopting the steepest descent rule which are as follows in the following equations (5) and (6)

$$W^n(k+1) = W^n(k) - \alpha s^n (G^{n-1})^T \quad (5)$$

$$b^n(k+1) = b^n(k) - \alpha s^n \quad (6)$$

where, α is the learning rate.

The BPNN model was developed in the software MATLAB and the training dataset which is 70% of the total dataset was utilized to develop the model and the remaining 30% of the model was adopted to assess the trained model.One of systems which can learn adaptively is DENFIS. This system can able to change dynamically every model to particular output for each input. If a relationship which is complex must be modelled, then DENFIS is trust worth approach. A experimented has conducted where DENFIS has given a model with 2 inputs to estimate ET for minimum and maximum temperatures. ANFIS is Adaptive Neural Fuzzy Inference Systems is similar software but Advaned version of it is DENFIS. Neuro-fuzzy models has been used by ANFIS for input and output to build a relationship called predictor-predictand. The drawback of ANFIS is because of stationary structure its arduous to adapt for the inputs which are new. In addition, ANFIS produces NF models which grows exponentially as per the numerous inputs given, which turns hard for solving high dimensional inputs. The DENFIS system uses rising clustering functions with many steps are taken to predict the cluster centers. This cluster data generates NF models. The massive noise in the data can be managed by the algorithms which are clustered in DENFIS makes the system model efficiently. There are two ways to develop the DENFIS system; Offline and online learning system. DENFIS develops a function, which is modified many times by Least-Square adopting method. If the process is offline, a better accuracy can be achieved using best learning algorithm for the dynamic evolution of NF models. (Jitter plot).In the study , the structure used is feed forward neural network and method is described as follows:This method mainly contains 2 phases. They are forward and backward computing.

In first case, which is forward computing, every layer has a weight matrix which has a connection with one behind layer and also to the upcoming layer. The weight matrix of the hidden layer is W_{ij} with f^1 as the activation function. Then the output layer also has the activation function f^2 and the weight matrix is W_{jm} Then the input vector of network is x belongs to R^{nx1} and the response vector which is output vector is y belongs to R^{nx1} .

The relation between them is written as follows

$$y_m = f^{(2)}\left\{\sum_{j=1}^{m}\left[f^{(1)}\left(\sum_{i=1}^{n}x_i W_{ij} + b_j\right)\right]W_{jm} + b_m\right\}$$

In this study, most used three transfer functions are Log-sigmoid, liner and tangent sigmoid and examined in experiments.Backward computing which is after forward computing, relay on the algorithms to settle the weights. This method of normalizing the weights to reduce the differences among desired and actual outputs is known as network learning or training. In a case where the errors are higher then the expected results, then the errors are repeated through and through backward in the network till minimized. In terms of ANN this stage is also known as algorithm of Back-propagation.

According the methods used to train the models of ANN(feed-forward neural networks), variety of back-propagation algorithms are generated and adopted. In a study, one of the algorithms of back-propagation called Levenberg-Marquadt, which is faster and second order nonlinear optimization method is used. This algorithm uses a newton method which is simplified and applied for adoption. The process of training should be considered a set of inevitable weights that reduce error (ep) for all samples in the group.Human brain's learning skills inspires for information processing methods of ANN. To establish the present patterns among input and output data sets, ANN requires samples to adopt accordingly. Also it is must examine the ability of ANN when given the input the DATA from outside of the sample given. To conduct entire process, including testing and training, the data we have should be randomized and divided accordingly. Approximately, the data for training the ANN is 70% and for testing only 30%, which is divided equally into sets for testing and validation.

To train the ANN, the data from 70% is used, which higher than the testing as ANN requires mode data to establish and understand the relationship between the data pairs of input and output updating the work weights of neural network systematically and BO algorithm. While training is going on, there is a chance for ANN to over-learn of over-fit the sample patterns given while training. This can be resulted for the low generalization of the network when given new data.

To over come the over fitting the ANN work, the data from Validation can also be used but indirectly So that this helps to reduce the errors during validation. Once the training is finished successfully then the testing of ANN start in second phase from the remaining data. The ANN has to predict the results using the training algorithms for rest of the input given from the data with as less as errors possible.

4. Results and Discussion

There are no norms for modifying the number of hidden neurons, a hit and trial approach were utilized for determining the optimum number of hidden neurons (Naik et al., 2014). In the upcoming table (1) the syntax for the various training functions and their corresponding algorithm type has been shown.

Table 1: BPNN model's training functions and their corresponding algorithm.

S.No	Type of Algorithm	Syntax of the training function
1	Scaled Conjugate Gradient	trainscg
2	Levenberg-Marquardt	trainlm
3	Fletcher-Reeves Conjugate Gradient	traincgf
4	BFGS quasi-Newton backpropagation	trainbfg
5	Gradient Descent with Momentum and Adaptive Learning Rate	traingdx
6	Bayesian Regularisation	trainbr
7	Polak-Ribiére Conjugate Gradient	traincgp
8	Conjugate Gradient with Powell/Beale Restarts	traincgb

The tuning parameters such as training algorithms, hidden transfer function, output transfer function and number of hidden neurons were determined by trial-and-error approach. The table (1) shows the utilization of various training algorithm that the BPNN model can adopt. The transfer functions such as "tansig" "logsig" and "purelin" were utilized.

According to the above-mentioned table, maximum of 2 hidden neurons were adopted in order to obtain the best result. The epochs are nothing but the iterations, which exposed that the number of iterations took place at lease mean squared error and the time taken to achieve the iteration with best possible coefficient of correlation value.

The above table 1 depicts that, the model number M4 and M6 with the training algorithm 'trainbgf' and 'trainbr' with the same transfer functions 'logsig' and 'tansig' were providing the best coefficient of correlation (R) value with minimal error.

The following table 2 depicts the neural architecture and the design parameters for the BP model.

Table 2: Neural structure and the design parameters of the BPNN model.

Trial Model No	Training Algorithm	Hidden transfer function	Output Transfer function	Neuronal architecture (input-hidden-output)	Epochs	Time (sec)	MSE	Validation	
								Training (R)	Testing (R)
M1	trainscg	tansig	logsig	3-2-1	300	5	0.000289	0.9968	0.9889
M2	trainlm	tansig	logsig	3-1-1	300	10	0.000223	0.9970	0.9894
M3	traincgf	purelin	logsig	3-1-1	19	1	0.000248	0.9968	0.9915
M4	trainbgf	logsig	Tansig	3-1-1	76	2	0.00008585	0.9988	0.9934
M5	traingdx	logsig	Logsig	3-1-1	140	3	0.0006883	0.9914	0.9769
M6	trainbr	logsig	Tansig	3-1-1	150	5	0.000062	0.9991	0.9945
M7	traincgp	purelin	Tansig	3-1-1	21	1	0.000714	0.9906	0.9884
M8	traincgb	tansig	Tansig	3-1-1	53	2	0.000128	0.9983	0.9926

. The performance of the model M4 and M6 can be depicted in the upcoming figure 3 and 4.

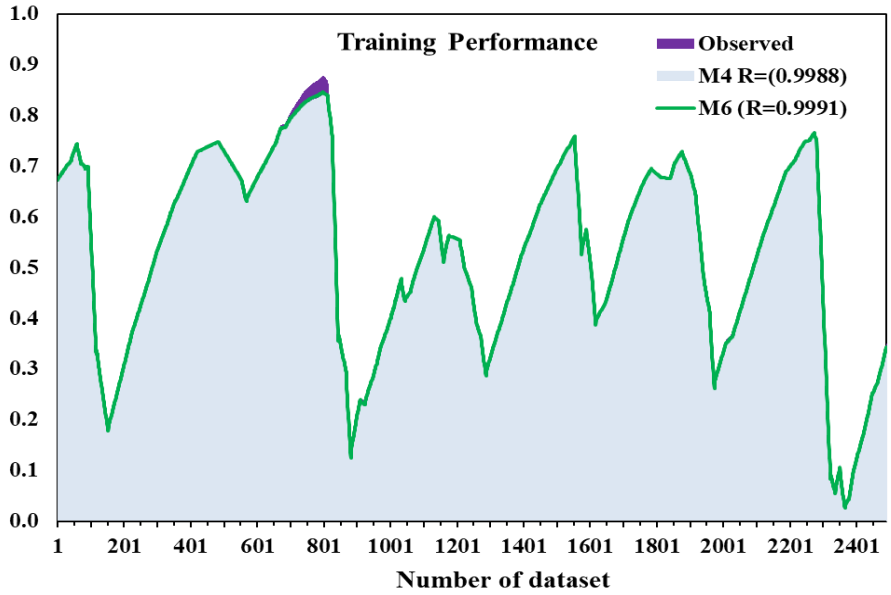


Figure 3: Training Performance of the BPNN model of M4 and M6.

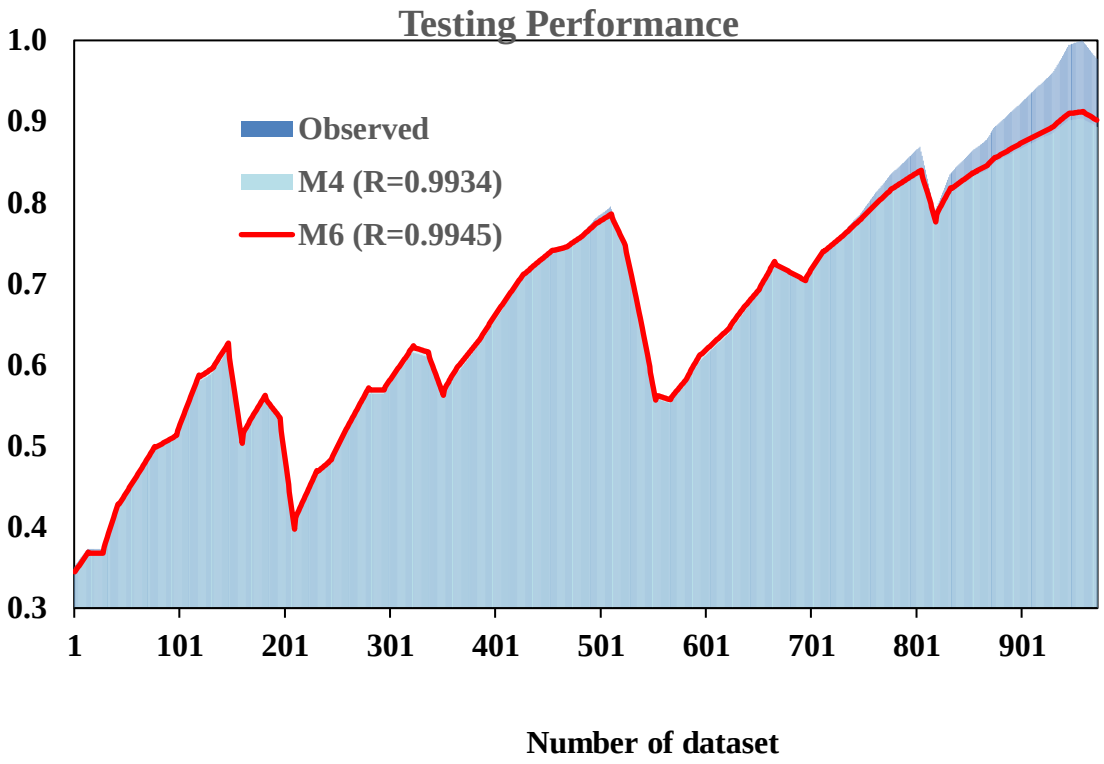


Figure 4: Testing Performance of the BPNN model of M4 and M6.

The above figure depicted that the models M4 and M6 performs well in determining the daily groundwater table, however the model M6 outperforms the other models. The upcoming figure (4) explains compares the potential of the various adopted training algorithm with respect to

their coefficient of correlation (R) and the time taken for forecasting the groundwater level. The figure (4) explains that, the model M6 outperforms other BPNN models.

There are other various statistical parameters such as Root mean squared error (RMSE), Mean absolute error (MAE), Coefficient of determination (R^2), Normalized mean bias error (NMBE), Variance account factor (VAF), Nash Sutcliffe efficiency (NS), Performance Index (PI) and adjusted determination coefficient (Adj.R^2) are utilized in order to assess the capability of the developed BPNN model

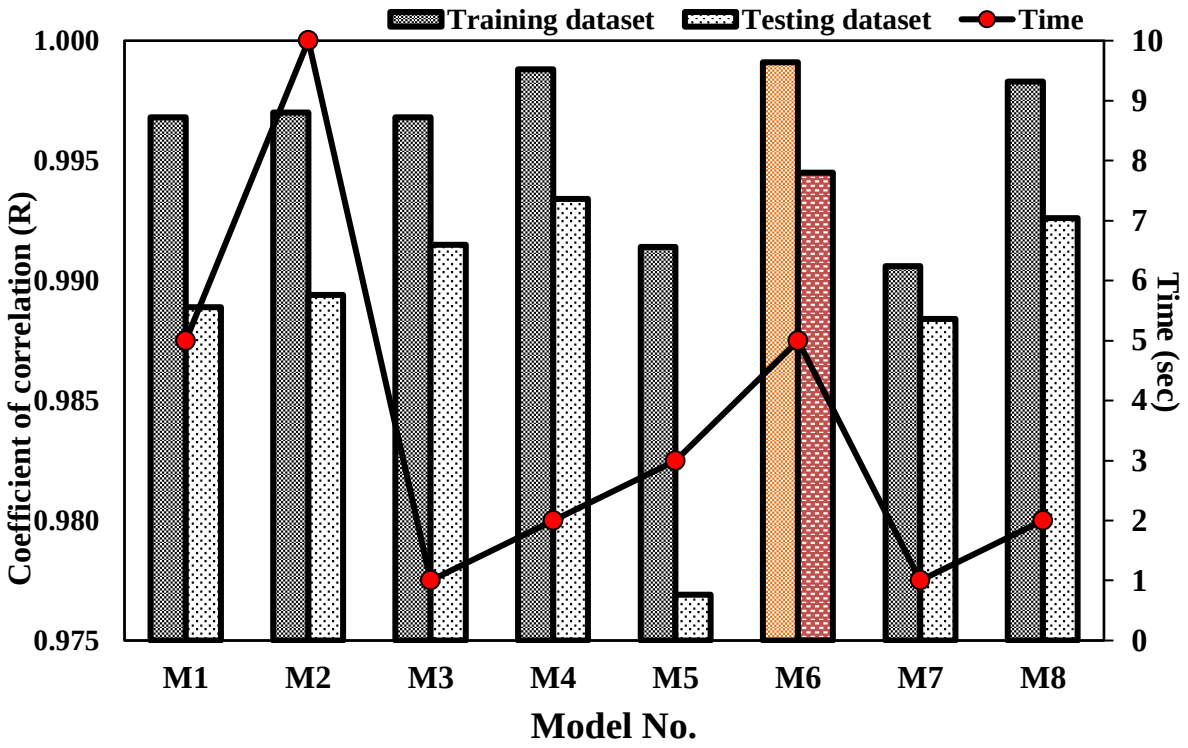


Figure 5.Comparison of performance of various training algorithms.

. The statistical values for the M4 and M6 models are depicted in the following table 3.

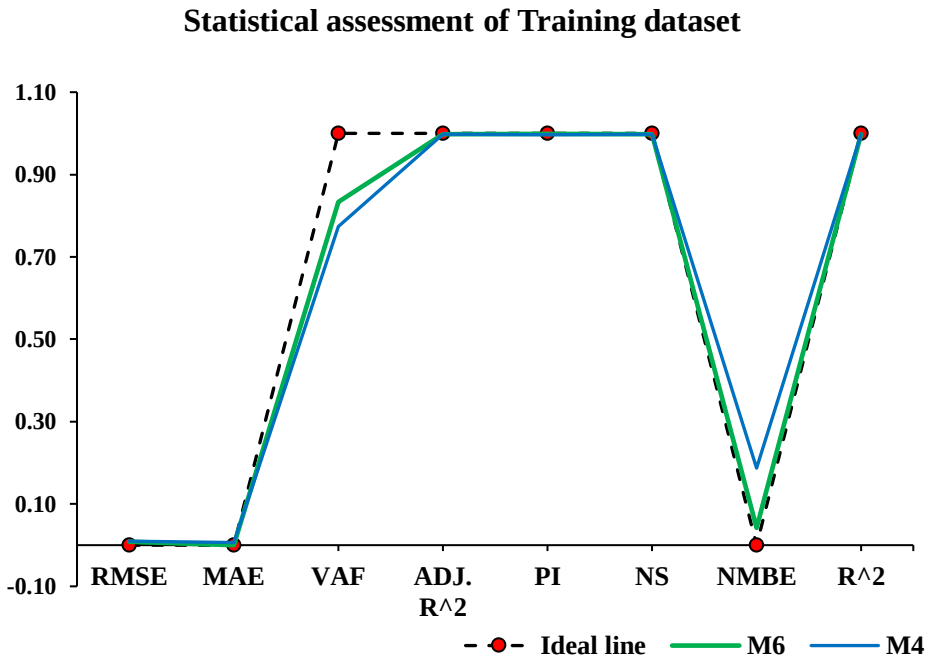
Table 3: Statistical assessment of the models M4 and M6.

STATISTICAL PARAMETER	**M4**		**M6**	
	Training	**Testing**	**Training**	**Testing**
RMSE	0.00927	0.02557	0.00786	0.00021
MAE	0.00636	0.01387	0.00000	0.00001
VAF	0.77380	-1.29526	0.83399	0.69525
ADJ. R^2	0.99760	0.98682	0.99820	0.98901
PI	0.99607	0.94829	0.99868	0.99575

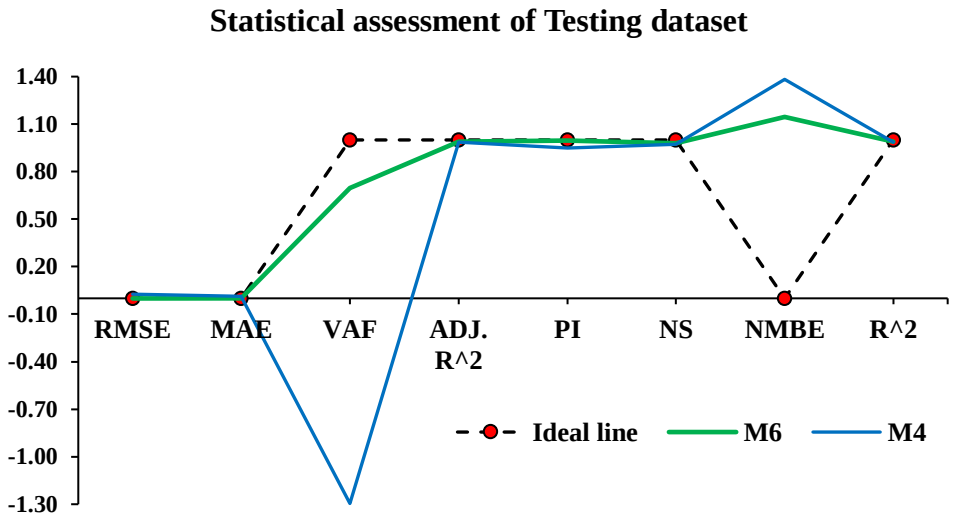
NS	0.99771	0.97360	0.99835	0.97873
NMBE	0.18683	1.38263	0.04245	1.14537
R^2	0.99760	0.98684	0.99820	0.98903

The above table 3 depicted that the errors such as RMSE and MAE for both M4 and M6 are less, however, the M6 model has less error than the M4 model. If the value of NMBE is positive then it represents the over prediction and vice versa as under prediction. From the above table 3, the value of NMBE for M6 model is more efficient than the M4 model.

Almost in most of the statistical parameters, the M6 model which has the training algorithm "trainbr" outperforms the other adopted training algorithms for predicting the daily groundwater table. The following figure (5a) and (5b) depicts the pictorial representation of the statistical assessment by comparing the M4 and M6 models.



6(a)



6(b)

Figure6(a) and 5(b) compares the statistical values of the developed M4 and M6 BPNN model.

5. Conclusion

The precise determination of the groundwater table can assist to guide the sustainable development and management of water usage. This study adopted one of the machine learning technique, BPNN in order to forecast the daily groundwater table variations for the specific location in Brazil. The capability of BPNN model by utilizing various training algorithm has been depicted in this study, however the method Bayesian regularization backpropagation with the syntax of "trainbr" with 1 hidden neuron and the transfer functions are logsig and tansig provided the best optimized output. The statistical performances of M6 model were compared with the other M4 model and hence the Bayesian regularization backpropagation method provided the best result. The adopted M6 model, when compared with the other models clearly depicted the accuracy and greater performance in every criterion.

References

1. Reinecke R, Wachholz A, Mehl S, Foglia L, Niemann C, Döll P. Importance of Spatial Resolution in Global Groundwater Modeling. Groundwater 2020;58:363–76.
2. Le Brocque AF, Kath J, Reardon-Smith K. Chronic groundwater decline: A multi-decadal analysis of groundwater trends under extreme climate cycles. J. Hydrol. 2018; 561: 976–86
3. Chang, F.-J., Chang, L.-C., Huang, C.-W., & Kao, I.-F. (2016). Prediction of monthly regional groundwater levels through hybrid soft-computing techniques. Journal of Hydrology, 541, 965–976.

4. Yadav, B., Gupta, P. K., Patidar, N., & Himanshu, S. K. (2019). Ensemble modelling framework for groundwater level prediction in urban areas of India. Science of The Total Environment, 135539.
5. Panahi, M., Sadhasivam, N., Reza Pourghasemi, H., Rezaie, F., & Lee, S. (2020). Spatial prediction of groundwater potential mapping based on convolutional neural network (CNN) and support vector regression (SVR). Journal of Hydrology, 125033.
6. Belkhiri, L., Tiri, A., & Mouni, L. (2020). Study of the spatial distribution of groundwater quality index using geostatistical models. Groundwater for Sustainable Development, 100473.
7. Ahmedbahaaaldin Ibrahem Ahmed Osman, Ali Najah Ahmed, Ming Fai Chow, Yuk Feng Huang, Ahmed El-Shafie, Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia, Ain Shams Engineering Journal, 2021,
8. Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986), "Learning Representations by BackPropagating Errors", Nature, 323, 6088, pp.533-536
9. Wang, H., Sánchez-Molina, J. A., Li, M., Berenguel, M., Yang, X. T., & Bienvenido, J. F. (2017). Leaf area index estimation for a greenhouse transpiration model using external climate conditions based on genetics algorithms, back-propagation neural networks and nonlinear autoregressive exogenous models. Agricultural Water Management, 183, 107–115.
10. Zhang, N., Ma, Y., & Zhang, Q. (2018). Prediction of sea ice evolution in Liaodong Bay based on a back-propagation neural network model. Cold Regions Science and Technology, 145, 65–75.
11. Lu, S., Li, Q., Bai, L., & Wang, R. (2019). Performance predictions of ground source heat pump system based on random forest and back propagation neural network models. Energy Conversion and Management, 197, 111864.
12. Zhou, S., Liu, N., Shen, C., Zhang, L., He, T., Yu, B., & Li, J. (2019). An adaptive Kalman filtering algorithm based on back-propagation (BP) neural network applied for simultaneously detection of exhaled CO and N2O. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 117332.
13. Zhang, H., Li, Y., & Zhang, H. (2019). Risk early warning safety model for sports events based on back propagation neural network machine learning. Safety Science, 118, 332–336.
14. Wang, G., Awad, O. I., Liu, S., Shuai, S., & Wang, Z. (2020). NOx emissions prediction based on mutual information and back propagation neural network using correlation quantitative analysis. Energy, 198, 117286.
15. Bo Wang, Bowen Chen, Gongqing Wang, Ru Li, Jiming Wen, Chuan Lu, Ruifeng Tian, Jian Deng, Back propagation (BP) neural network prediction and chaotic characteristics analysis of free falling liquid film fluctuation on corrugated plate wall, Annals of Nuclear Energy, Volume 148, 2020, 107711.
16. Taheri, K.; Shahabi, H.; Chapi, K.; Shirzadi, A.; Gutiérrez, F.; Khosravi, K. Sinkhole susceptibility mapping:A comparison between bayes-based machine learning algorithms. Land Degrad. Dev. 2019, 30, 730–745.

17. Pham, B.T.; Prakash, I.; Khosravi, K.; Chapi, K.; Trinh, P.T.; Ngo, T.Q.; Hosseini, S.V.; Bui, D.T. A comparison of support vector machines and bayesian algorithms for landslide susceptibility modelling. Geocarto Int. 2019, 34, 1385–1407.
18. Chen, W.; Hong, H.; Panahi, M.; Shahabi, H.; Wang, Y.; Shirzadi, A.; Pirasteh, S.; Alesheikh, A.A.; Khosravi, K.; Panahi, S. Spatial prediction of landslide susceptibility using gis-based data mining techniques of anfis with whale optimization algorithm (woa) and grey wolf optimizer (gwo). Appl. Sci. 2019, 9, 3755.
19. Khosravi, K.; Daggupati, P.; Alami, M.T.; Awadh, S.M.; Ghareb, M.I.; Panahi, M.; Pham, B.T.; Rezaie, F.; Qi, C.; Yaseen, Z.M. Meteorological data mining and hybrid data-intelligence models for reference evaporation simulation: A case study in iraq. Comput. Electron. Agric. 2019, 167, 105041.
20. Khozani, Z.S.; Khosravi, K.; Pham, B.T.; Kløve, B.; Mohtar, W.; Melini, W.H.; Yaseen, Z.M. Determination of compound channel apparent shear stress: Application of novel data mining models. J. Hydroinform. 2019, 21, 798–811.
21. Feng, Y., Liu, Y.-Z., Wang, X., He, Z.-X., Hung, T.-C., Wang, Q., & Xi, H. (2020). Performance prediction and optimization of an organic Rankine cycle (ORC) for waste heat recovery using back propagation neural network. Energy Conversion and Management, 226, 113552.
22. Jana, G. C., Swetapadma, A., & Pattnaik, P. K. (2018). Enhancing the performance of motor imagery classification to design a robust brain computer interface using feed forward back-propagation neural network. Ain Shams Engineering Journal, 2871-2878.
23. Naik, A.R., Shafi, P.M., Kosbatwar, S.P., 2014. Weather prediction using error minimization algorithm on feedforward artificial neural network. In: Mohapatra, D.P., Patnaik, S. (Eds.), Intelligent Computing, Networking, and Informatics, Advances in Intelligent Systems and Computing. Springer, New Delhi, India.

AUTHORS PROFILE



Varna Vishakar V currently working as Assistant professor in KIET Group of Institution having Masters in Environmental Specialization from Anna University, Chennai, published 6 International Paper publications in Scopus and other reputed journals. Experienced Chairperson with a demonstrated history of working in the higher education industry. Skilled in Civil Engineering, Environmental Engineering, Environmental Conservation Studies, Impact Assessment, Water Quality Monitoring, Sustainable Development, AutoCAD, Technical Presentation, Project Managing, and Project Coordination.



Ayush Jain, currently working as Assistant professor in KIET Group of Institution has masters in Alternate Hydro Energy System from Indian institute of technology Roorkee and completed his graduation in Civil Engineering from Rajasthan Technical University published 2 papers related to the Land use and land cover dynamics.



Zohaib Ahmed Khan is an Assistant Professor in the Department of Civil Engineering at KIET Group of Institutions, Delhi-NCR, Ghaziabad, Completed M.Tech in Hydraulics and Water Resources Engineering from Delhi Technological University.



Ayush kumar currently working an Assistant Professor in the Department of Civil Engineering at KIET Group of Institutions, Delhi-NCR, Ghaziabad, Completed M.Tech in Environment Science & Engineering from Jamia Millia Islamia, New Delhi.

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Ethical Framework for Iot in People Analytics: Risks and Opportunities

[1]Mr. Nishant Gaur
Dr. Vikas Gupta

Submitted: 06/06/2022 **Accepted**: 10/09/2022

About the Authors:

Mr. Nishant Gaur is research scholar in Delhi school of Management, DTU. His area of interests are Knowledge management, ethics, people analytics, organizational culture and Organizational behavior.

Dr. Vikas Gupta is Assistant Professor in Delhi school of Management,, DTU. He is having 20 years of experience. His area of interests include Knowledge Management, Knowledge management practices, Innovation Management, Business Process engineering.

Abstract

Purpose: This research studies the ethical perspective concerning internet of things (IoT) placement in People Analytics. To provide researchers and professionals with ethical framework for IoT (internet of things) in people analytics. .Also to identify associated risk and opportunities.

Design/methodology/approach: Initially, the applicability of Chuck Huff's original Personality, Integration of morality, Moral Ecology and skills Model (PRIMES) is studied from context of IoT in people analytics. Secondly consideration of ethical issues in addition to PRIMES model are proposed based on limited scope of PRIMES in IoT and people analytics.

Findings: The original PRIMES Model can be utilized in initial stages as ethical guiding framework for individuals employed in personnel department but it lacks coverage of ethical issues from perspective of IoT in people analytics. To address the ethical dimensions from IoT in people analytics viewpoint, additional ethical issues are addressed.

Research limitations: The novel ethical framework for IoT in people analytics required further authentication and validation along with empirical testing in continuously emerging IoT and people analytics ecosystem.

Implications: Considering the paucity of ethical frameworks in emerging area of IoT in people analytics, this study provides the ethical model in the area of IoT in people analytics for the researchers and practitioners. This framework can further be tested and used practically and can also be considered for theoretical development.

Originality/Value: There is apparent deficiency of ethical norms in area of IoT in people analytics, this study contributes to the area by providing extended PRIMES Model as a preliminary ethical framework for IoT in people analytics.

Keywords: Ethics, Internet of Things, People Analytics, PRIMES, Human Resource Management.

Paper type: Conceptual

[1] Research Scholar
Delhi School of Management,
Delhi Technological University, Delhi
gaurnishant2006@gmail.com

Delhi School of Management,
Delhi Technological University, Delhi
E mail id: vikasguptadtu@gmail.com
ORCID ID: 0000-0003-2495-3669

1. Introduction

The Internet of Things (IoT) is novel and rapidly evolving platform in era of knowledge economy which enables its user to observe, create, store, interpret, moderate and use data acquired through physical objects embodied with sensors, software and other technologies that interact and exchanges data with other entities over internet and other communication networks (Vermanen, Rantanen and Ville,2021). At the same time innovation in human resource management (HRM) through people analytics is an emerging area which is making human resource (HR) system as data driven through studying people, process and techniques at work so as to elevate the human resource system for sustainable development. People analytics is gaining importance with advancement in digital and data science methods to explore, investigate and visualize information (Tursunbayeva, Pagliari and Lauro and Antonelli, 2021). People analytics strategically optimizes the employees, teams, departments and workforce through the application of IoT in human resource management (Fitz-Enz and Mattox, 2014). IoT may be applied in people analytics at individual level, departmental level and enterprise level (Edwards and Edwards, 2016). IoT captures employee's performance, skills, and abilities throughout life cycle of employee from acquisition to separation for effective outcomes in people analytics. This data can be used to identify employee's potential, assess performance data, designing personalized training programmes, forecast future performances and infer employee satisfaction, tracking employee wellness (Nunn, 2018). Personal analytics also extends beyond work life by keeping an eye on employees personal emails, usage and interaction with digital devices and apps, tracking employee's activities on social media platforms. People analytics using IoT acquires employees personal data related to their purchases, location, health, activities and social life (Ajunwa, I., Crawford, K. and Schultz, J., 2017). Employee data captured through IoT devices are used to design algorithms for people analytics to shape their behavior on the basis of recorded pattern of their existing behavior through rewards, appreciation, incentives and other techniques (Cardador, M.T., Northcraft, G.B. and Whicker, J., 2017). The use of IoT in people analytics is encouraging from organizational perspective but present the risks for employee's privacy, autonomy and confidentiality. The IoT in people analytics may unmask the hidden unethical practices which may include fraudulent expense claims, gender bias and intellectual property theft (Holeman, I., Cookson, T.P. and Pagliari, C., 2016).

The number of successful implementation and execution of IoT devices is increasing and business of IoT devices is expanding with the cost of IoT devices declining (Vermanen and Harakke,2019). With decreasing costs of IoT devices, the ethical concern in the area is increasing. As IoT devices offers several solutions, at the same time they come with many associated risks also (Hoyer, V., Janner, T., Mayer, P., Raus, M. and Schroth, C., 2006). There is a strong need to explore and identify the practices that can facilitate the ethical consideration while execution and implementation of IoT devices specifically in people analytics. Thus in this paper, we have developed an ethical framework for IoT devices in people analytics.

2. Background: Internet Of Things For People Analytics

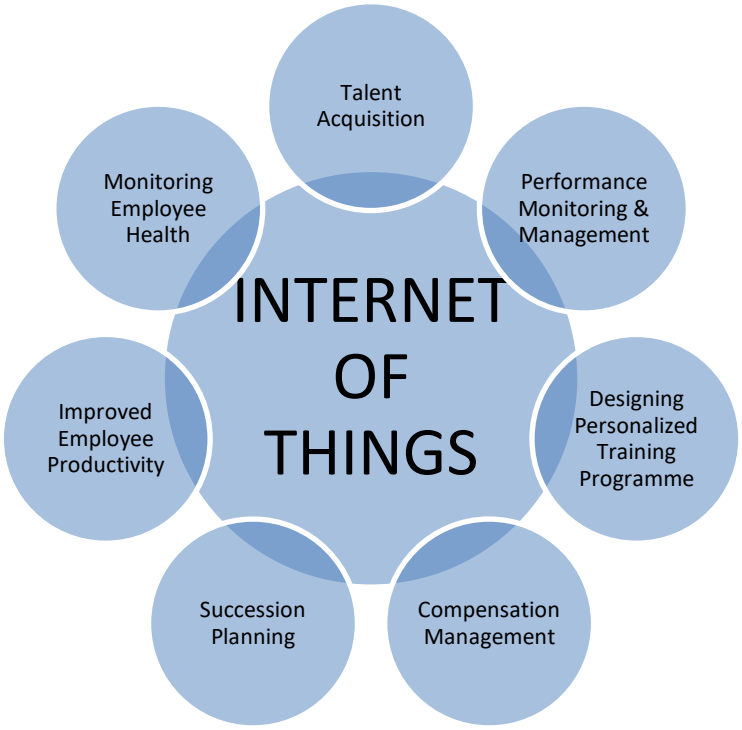
Organizations are progressing through development, advancement and expansion in technological infrastructure by implementing artificial intelligence, robotics, machine learning, data sciences, IoT and people analytics (Gaur Bhawna, Shukla Kumar Vinod and Verma

Amit, 2019). IoT is a nexus of physical objects that are implanted with sensors, software, and other technologies which are connected to internet for communicating and exchanging information with a goal to create smart organizations and smart homes (Wortmann, F. and Fleuchter, K., 2015). IoT binds end users with several physical objects embedded with sensors, software, tracking devices for collecting the data of interest (Gubbi, J., Buyya, R., Marusic, S. and Palaniswami, M., 2013). The data is gathered through these objects and delivered to end users (Jin, J., Gubbi, J., Marusic, S. and Palaniswami, M. ,2014). Through IoT applications companies perform on high accuracy and with utmost efficiency. IoT is most leveraged by people analytics for enhanced business decisions in context to procurement, development, performance monitoring, compensation, optimum utilization, behavior shaping and retention of talent in organization. IoT has radically transformed the traditional HR practices into people analytics through better use and further application of people data. IoT device collect the real time data from employees through sensors and transmit the data to the system through communication network enabled with internet. IoT have majorly reformed Human resource management through people analytics. IoT is facilitating and making HR activities convenient through better recruitment, performance monitoring, tracking employee wellness, attitude measurement, behavior tracking which is leading to accurate insight of work done. People analytics helps HR professionals for optimum utilization of big data for decision making .People Analytics utilize and harvest historical data so as to reap its benefit in future decision making and strategy building through designing suitable algorithms. People analytics have transformed traditional HRM practices and resulted in better workplace learning, motivating employees, identifying potential for future tasks, assessing candidates through regular performance tracking and reducing biasness.

IoT is used by HR managers to develop people analytics for operational excellence in various traditional HR practices. Figure 1 depicts contribution of IoT towards talent acquisition and on boarding, performance monitoring & management, designing personalize training programme, succession planning, compensation management, improved employee productivity, monitoring employee health. IoT tools are used for acquisition of personnel and to minimize the perceptual biasness in decision making related to recruitment and selection. IoT has automated and optimized the process of recruitment resulting in improved quality of hiring and reduced personal biasness. IoT devices monitor the performance of employees on real time basis and data accumulated through these IoT devices is used to develop performance management system through analytics. On the basis of performance monitoring, management and evaluation through IoT devices, personalized training programmes are developed. Data captured by IoT devices enables the HR manager to decide the candidates who can occupy senior positions in organizational hierarchy for better succession planning. IoT devices help the HR department in managing payroll and other expenses like transportation bills and other expenses. IoT devices are attached to employees as well as their systems or equipments so as to measure the employees experience at work and to develop strategies to improve their productivity. The companies monitor employ wellness through providing them fitness trackers which tracks their physical movement, calorie count, stress level which are related to health issues like heart disease, obesity and diabetes.

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

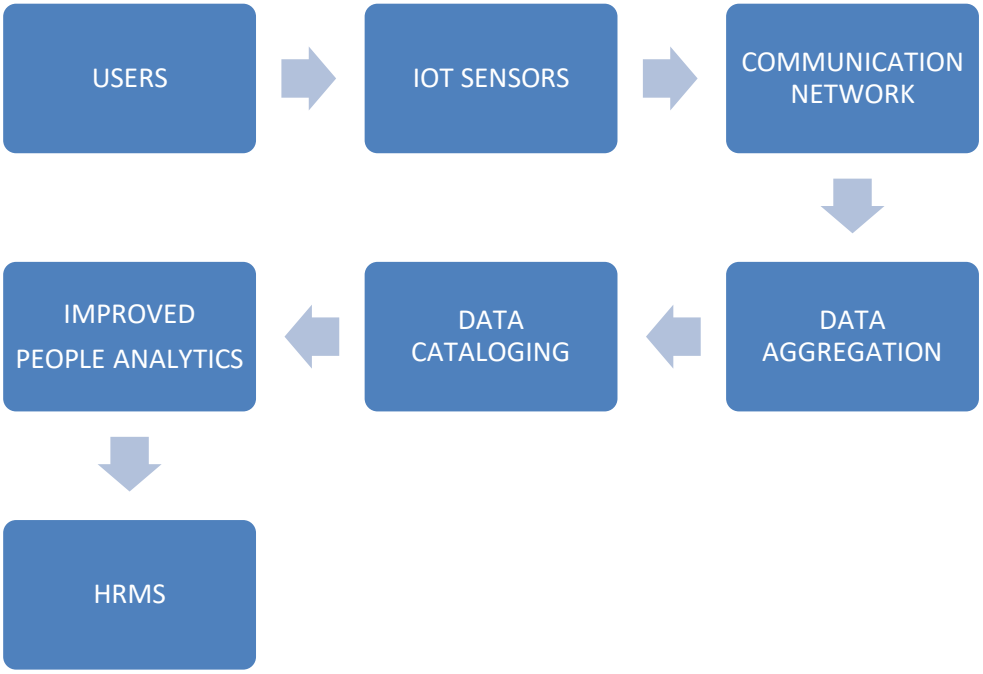
Figure 1. Connectivity of Basic HRM Functions with IoT Devices



Source: Compiled by authors

Figure 2 demonstrates the process of data collection through IoT devices in people analytics. The various IoT devices used by the HR department are equipped with various sensors like location tracking, pulse rate monitoring, and thermo sensor to measure body temperature. This data can facilitate real time decision. Data is collected on continuous basis with the help of sensors which is further disseminated through organization network. Further to this stage, data is summarized in an organized and systematic manner for statistical analysis and investigation. This stage is referred to as data aggregation. After data aggregation, the data is fragmented and stored in local databases and cloud for maximum utilization. This process leads to betterpeople analytics and HRMS

Figure 2. Process of data collection with IoT devices



Source: Bhawna Gaur, Vinod Kumar Shukla and Amit Verma (2019)

3. Ethical Challenges Of Internet Of Things In People Analytics

IoT devices and solutions bring many benefits and simultaneously bring various risks associated with privacy and safety of employees (Conti, M., Dehghantanha, A., Franke, K. and Watson, S., 2018). The safety and security of every IoT device which includes sensor, software and information unit are compromised with excessive use, declining prices and vulnerabilities leading from excessive growth of IoT market where security measures fail to cope up with all associated risks (Giaretta, A., Balasubramaniam, S. and Conti, M., 2016). These concerns can have symbolic consecutions from individual perspective as IoT devices can also collect employee's private and personal data (Lee and Lee, 2015). IoT devices can collect personal chat conversation, location, movement, shopping behavior, purchasing preferences, health conditions. As the IoT device keeps on monitoring the location, IoT device can also monitor the places visited by the person, whom the person is meeting, where person is travelling along with the current and live location of that person being tracked. The IoT devices in people analytics not only monitor employee's performance, assess training needs, monitors wellness but also invades into employees privacy. People analytics is a special case in handling of both technology and human resource data. The most relevant attribute of people analytics is people-centered style of knowledge management. This attribute of people analytics influence the ability of people from HRM department to tackle ethical queries in their HR data, IoT devices and information systems. However it is very much debated that the consecutions of increased privacy protection may not be totally positive, as this could delimit the uses and benefits that could be attained from IoT devices and data collected through it. IoT infrastructure possesses heterogeneous setup with hardware, software, sensor, network which creates much complexity in managing privacy and security issues. There is a need to identify,

explore and establish practices and policies that guides ethical implementation of IoT in people analytics. To intensively address and comprehend the need of ethics in IoT driven people analytics, a comprehensive ethical framework is needed.

4. Research Approach

This section constitutes the exploration and identification of research literature focused towards ethical framework for IoT in people analytics. While the main area of inclination was literature in ethics and IoT from HRM perspective. The literature was explored to allow for disciplinary breadth. The searches were carried out in two multidisciplinary databases – Web of Science Core Collection (WoS) and Scopus in March 2022. Google scholar was utilized to ensure all-inclusiveness of results returned from the databases.

The search combined the terms ethics with IoT and people analytics as well as human resource management. . An example Boolean search would be ("ethics" or "ethical framework" or "model of ethics") and (Internet of Things) and (people analytics or human resource management). The result of searches in WoS and Scopus were 23 and 31 with total of 54 results. Total 41 results were attained after elimination of duplicates. Four more results were eliminated as they were in language other than English. Two more results were removed as they were revealing the similar study as other articles in list. Finally only studies with empirical work and/ or studies with conceptual framework studying ethics in IoT driven people analytics or HRM was included. While satisfying this criteria nine more papers were excluded. At last after the thorough examination, 21 papers were finalized and were further analyzed and summarized.

After conducting this process, HRM functions facilitated through IoT devices, data collection through IoT devices, model of ethics (PRIMES) and ethical issues were identified. Theoretical and empirical findings from previous literature have been reviewed to support the relevance of ethical framework for IoT in people analytics.

5. Ethical Model (Primes) From Hrm Pperspective

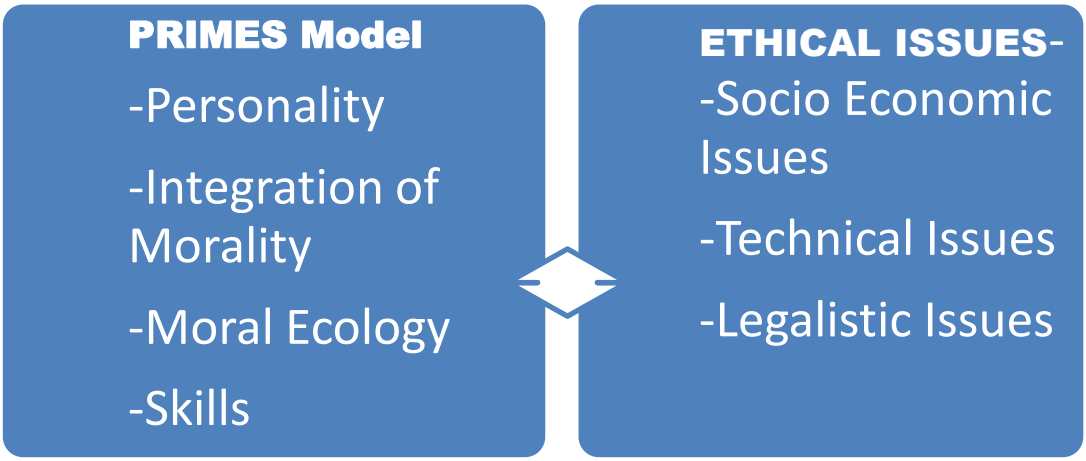
Ethicality in information age is of vast importance as information systems have high degree of invasion in individual's privacy and ethics in technology related matter is dependent on nature, use and potential of technology. Since IoT in people analytics is majorly driven by personnel's of HR department, so the model of ethics under considering here is PRIMES which focuses on ethics from personality, skill and morality of surrounding people viewpoint. This model is being utilized from perspective of its application to the employees of HR department. As employees or teams from HR department are the one who are totally involved in people analytics and deployment of IoT for achievement of best and effective results from people analytics. In people analytics driven through IoT, employees from HR department are the main drivers for ethicality or un-ethicality in the organization. Thus the understanding of ethics by a individual or a team, skills to handle ethical dilemmas, commitment towards pursuing ethical deeds, orientation of entire team towards morality lays down the foundation of ethical framework in IoT driven people analytics. PRIMES Model constitutes personality, Integration of morality, moral ecology and skills (Chuck Huff, 2010). Personality in PRIMES model indicates how individual carry out work with ethical orientation and is totally independent from external factors like possession of ethical behavior by team, department, organization. Personality trait of ethics is clearly visible in the action of people at work while carrying out their routine duties and responsibilities. Personality component of PRIMES model is formed from various factors

like family background and past experiences**.** The second component of PRIMES model is Integration of morality. Integration of morality refers to individual's commitment towards ethics and moral action. Integration of morality constitutes intensity with which individual act morally and his persistence towards moral behavior. The third component of PRIMES Model is moral ecology. Moral ecology refers to morality and moral action of people in surrounding who are part of same team, group and department. Surrounding constitutes of both moral and immoral action which can both support and hamper moral action of individual **(**Huff, C.W., & Barnard, L.,2008**).** The last component of PRIMES model is skill. Skill is the ability of individual to carry out moral actions at personal level as well as in team. Skill set support an individual in ethical dilemma encounters. The use of IoT in people analytics present risks for employee's privacy, autonomy and confidentiality. And the application of the ethical principles is totally dependent on the ethical orientation of individual, team and department which can be reflected through their personality, moral commitment, moral ecology and skill set. So PRIMES Model is best at ground level to establish ethical foundation for IoT driven people analytics.

6. Expanded Primes Model

PRIMES model provide with basic understanding of ethics for individual, team and department but fails to deliver comprehensive framework for IoT driven people analytics. As the IoT in people analytics unmask the hidden unethical practices which may include retrenchment or downsizing issues as expertise of employees is already captured through IOT devices, technical issues where devices are capturing the personal and private information of employees, intellectual property theft where individuals personal knowledge is claimed as organizations knowledge. (Land, F., & Urooj, A.,2003) explored three pivot dimensions to ethical issues which include socio economic issue, technical issue and legalistic issue. The organization always has a hidden objective behind implementing IoT in people analytics. The underlying motivation is to capture the expertise of employee through IoT devices in knowledge repositories and expert system (Bryant, 2006). Based on this agenda organization goes for downsizing as employees become less valuable. There are technical teams or outsourcing organizations that provide support in IoT implementation and execution. These third party companies or outsourcing firms are also involved in data theft or capturing of personal data of employees while designing installing and implementation the system. Many times the HR team can act as a whistle blower by being vigilant during design process, implementation and execution of IoT in people analytics. Also the data captured by IoT devices is claimed as organizations asset. This is prone to legalistic issues because of ownership conflict between the employee and the company. These ethical issues should also be embedded with the PRIMES model to provide a comprehensive ethical framework for IoT driven people analytics.

Figure 3. Ethical framework integrating PRIMES Model and Ethical Issues



Source: Compiled by authors

7. Risk And Opportunities

This section presents the associated risks and opportunities for employees and organizations in ethics for IoT driven people analytics. There are various risks for employees from IoT and people analytics which includes biasness and discrimination from collected data, psychological profiling, threat to privacy through tracking & surveillance and behavior shaping. Personal analytics based on IoT devices are designed with motive of enabling HR related decisions based on data instead of biased human perceptions and flawed human reasoning. But the IoT based people analytics is also designed under the supervision of same HR team so the biased behavior and potential for prejudice still remains same and is encrypted in algorithms (Dastin, 2018). As the recruitment portal may be designed discriminating female applicants leading to biasness towards female gender. It can also be incorporated for race, age, disability, sexual orientation leading to biasness and discrimination. In people analytics while conducting employee selection, personality tests are conducted through specially designed algorithms. These kinds of tests overlook moral character, cultural differences and perform psychological profiling (Kirke, 2019). IoT driven people analytics track and monitors employee in workplace as well as in personal lives through mobile phones or IoT devices like smart watch, systems given to employees from companies (Guenole, N., Feinzig, S. and Green, D., 2018),. The data captured through IoT devices on performance patterns of employees along with emails and other responses are utilized to design algorithms. These algorithms further works to transmit personalized messages to shape employee behavior to achieve individual's work related objectives as well as team and organizational objectives (High, 2019). But at the same time it lacks transparency as whether employees know their emails and other responses are being observed and tracked along with performance related data to frame the messages and repeated reminders which they receive resulting in their behavior shaping.

Ethics possess a point of risk to organizations for IoT driven people analytics (Peterson 2018). IoT and people analytics are relatively new so organizations lack appropriate ethical and legal knowledge to tackle prevailing situations. IoT driven people analytics projects fails as they compromise with privacy and security issues. Also existing risk management strategies are unsuccessful because they are unable to recognize indicators of failure in this new domain of IoT driven people analytics

(Clavard and Jeske, 2018). Employees also lack the trust in IoT driven people analytics as they believe employers are storing their expertise along with their personal and sensitive data (Pease 2018). Thus employees play smart and also game the system when they are already aware that they are under observation (Jacobs, 2017). The organizations also keep delaying their IoT driven people analytics projects due to uncertainty over there regulatory compliances. One of the ignored aspects in IoT driven people analytics is that design and implementation of IoT and people analytics is left over to technology outsourcing partners whereas it should be carried out under the strict observation of HRM team. Employees are criticizing and revolting against perceived misuses of data collected and stored through IoT and people analytics leading to failure of these projects (Helmore, 2019).

The IoT in people analytics deliver many benefits to the employee and organization. These benefits can be fully harnessed and further opportunities can be tapped by developing ethical framework for IoT driven people analytics. IoT and people analytics perform the major operative functions of HRM which includes employee acquisition and retention, assessing candidates through regular performance monitoring, performance management, designing personalize training programme, compensation management, tracking employee wellness and monitoring employee health, attitude measurement, behavior tracking for optimum job performance, identifying potential for future tasks leading to succession planning.

The risks associated with IoT and people analytics can be transformed into opportunities by devising an ethical mechanism to tackle all prevailing risks. The organizations should disseminate the purpose and benefits of implementing IoT and people analytics throughout the organizations which may lead to transparency and fairness amongst the employees in organizations. Transparency and fairness may deliver optimum results from IoT and people analytics (Manyika, 2019). The IoT and people analytics should completely adhere to legislation in order to avoid any allegation from employees and to avoid any kind of protest for the same (Kim, 2017). This may result in industrial peace and smooth execution of IoT and people analytics. The ethical guidelines and charter in regard to IoT and people analytics may prevent any conflict amongst employee and organization (Peterson, 2018). The employees should be made aware and sensitized about data tracking, data storage, data ownership permitted in their jurisdiction (Jones, 2018). Also, the people analytics team should take complete control over IoT implementation in people analytics rather than allowing technology outsourcing firms to independently design, execute and implement IoT in people analytics (Agarwal, D., Bersin, J., Lahiri, G., Schwartz, J. and Volini, E.,2018). The legal compliance officers should be consulted for regulation and polices in relation to IoT as people analytics team cannot know everything about data security and privacy and legal requirements (Green, 2018).

8. Conclusions And Implications

The curiosity of organizations in implementing information technology based infrastructure at organization level and specific department level has risen at an exponential rate. This era is to be marked as knowledge era, with steady rise in information technology infrastructure provided by IoT for HR analytics, marketing analytics and other related fields of management. This has pursued government, corporate and researchers to establish and manifest ethical principles, guidelines and frameworks. This study primarily develops a comprehensive framework of ethics for IoT driven people analytics. The framework is expanded from the PRIMES model which

provides ethical ground for the related employees but lacks ethical coverage from IoT and people analytics perspective. The proposed framework integrates PRIMES model with three ethical concerns in relation to IoT and people analytics. The ethical deliberation for IoT and people analytics has received less attention being a new domain area. While exploring academic literature very limited information was available for ethics in IoT and people analytics. Although blog posts, discussion papers and subjective case reports were available on ethics in IoT and people analytics. But there was scarcity of conceptual models and empirical studies on ethics in IoT and people analytics. After analyzing all the ethical concerns raised by IoT and people analytics, organization may face serious repercussions and losses regarding employee trust and company image. IoT devices rarely limit unethical malpractices as data can be collected and stored anywhere and anytime despite of time and location. The existing law provides limited protection against security and privacy risks imposed by IoT and people analytics. The ethical responsibility primarily lies on individual's and group's moral standards. With the same belief this paper has utilized PRIMES model as basic approach towards studying and implementing ethics in IoT and people analytics as PRIMES cover individuals orientation towards ethics, commitment towards ethics, and ethics of people in surrounding environment and skills to navigate through ethical dilemmas. The socio-economic issues, technical issues and legal issues are also incorporated along with PRIMES to provide a comprehensive ethical framework to deal with ethical concerns in IoT and people analytics. The IoT use in people analytics is growing rapidly, so the need and relevance of ethical framework in the field is much required.

The study has contribution towards literature and also has practical contributions as it can be utilized by companies and people analytics team to yield better results from IoT and people analytics. The study has a relevant theoretical contribution as it conceptualizes ethical framework for IoT and people analytics. The expanded PRIMES model provides a platform for studying ethical issues further. This framework from practical viewpoint provides practicing managers to apply it in their people analytics project which are IoT diven. We suggest further this framework can be tested empirically and more dimensions can be added to the framework paving the way for future research.

References

[1]. Agarwal, D., Bersin, J., Lahiri, G., Schwartz, J. and Volini, E. (2018), "People data: how far is too far?, available at: https://www2.deloitte.com/za/en/pages/human-capital/articles/people-data-how-far is-too-far.html

[2]. Ajunwa, I., Crawford, K. and Schultz, J. (2017), "Limitless worker surveillance", California Law Review, Vol. 105 No. 3, pp. 735-776.

[3]. Bryant. (2006). Knowledge management: The ethics of the Agora or the mechanisms of the market? In the proceedings of 39th Annual Hawaii International Conference on System Sciences, January 4-7, 2006, Computer Society Press.

[4]. Calvard, T.S. and Jeske, D. (2018), "Developing human resource data risk management in the age of big data", International Journal of Information Management, Vol. 43, pp. 159-164.

[5]. Cardador, M.T., Northcraft, G.B. and Whicker, J. (2017), "A theory of work gamification: something old, something new, something borrowed, something cool?", Human Resource Management Review, Vol. 27 No. 2, pp. 353-365.

[6]. Conti, M., Dehghantanha, A., Franke, K. and Watson, S. (2018), "Internet of Things security and forensics: challenges and opportunities", Future Generation Computer Systems, Vol. 78 No. 2, pp. 544-546, doi: 10.1016/j.future.2017.07.060.

[7]. Dastin, J. (2018), "Amazon scraps secret AI recruiting tool that showed bias against women", Reuters, 9 October, available at: https://www.reuters.com/article/us-amazon-com-jobs-automation insight-idUSKCN1MK08G

[8]. Edwards, M.R. and Edwards, K. (2016), Predictive HR Analytics: Mastering the HR Metric, Kogan Page, New York, NY.

[9]. Fitz-Enz, Jac and Mattox, J.R. II (2014), Predictive Analytics for Human Resources, John Wiley & Sons, Hoboken, New Jersey.

[10]. Gaur Bhawna, Shukla Kumar Vinod and Verma Amit (2019), "Strengthening People Analytics through Wearable IOT Device for Real-Time Data Collection", 2019 International Conference on Automation, Computational and Technology Management (ICACTM) Amity University, 978-1-5386-8010-0/19/$31.00 ©2019 IEEE.

[11]. Giaretta, A., Balasubramaniam, S. and Conti, M. (2016), "Security vulnerabilities and counter measures for target localisation in bio-nano things communication networks", IEEE Transactions on Information Forensics and Security, Vol. 11 No. 4, pp. 665-676, doi: 10.1109/ TIFS.2015.2505632.

[12]. Green, D. and Chidambaram, A. (2018), The Role of the People Analytics Leader - Part 2: Creating Organisational Culture and Shaping the Future, myHRfuture, 25 February, available at: https:// www.myhrfuture.com/blog/2018/2/25/the-role-of-the-people-analytics-leader-part-2-creating organisational-culture-shaping-the-future.

[13]. Gubbi, J., Buyya, R., Marusic, S. and Palaniswami, M. (2013), "Internet of Things (IoT): a vision, architectural elements, and future directions", Future Generation Computer Systems, Vol. 29 No. 7, pp. 1645-1660, doi: 10.1016/j.future.2013.01.010.

[14]. Guenole, N., Feinzig, S. and Green, D. (2018), "The grey area: ethical dilemmas in HR analytics. Perspectives from the global workforce", IBM, available at: https://www.ibm.com/watson/ talent/talent-management-institute/ethical-dilemmas-hr-analytics/hr-ethical-dilemmas.pdf.

[15]. Helmore, E. (2019), "Hundreds of Google employees urge company to resist support for Ice", The Guardian, 16 August, available at: https://www.theguardian.com/technology/2019/aug/16/ hundreds-of-google-employees-urge-company-to-resist-support-for-ice#:~:text=Hundreds% 20of%20Google%20employees%20urge%20company%20to%20resist%20support%20for% 20Ice, This%20article%20is&text=A%20group%20of%20employees%20called,Immigration

[16]. %20and%20Customs%20Enforcement%20contract High, P. (2019), "Former Google HR chief Laszlo Bock aims to revolutionize people management with Humu", Forbes, 9 September, available at: https://www.forbes.com/sites/peterhigh/2019/09/09/former-google-hr-chief-laszlo- bock-aims-to-revolutionize-people-management-with humu/

[17]. Hoyer, V., Janner, T., Mayer, P., Raus, M. and Schroth, C. (2006), "Small and medium enterprise's benefits of next generation e-business platforms", The Business Review, Cambridge, Vol. 10 No. 2, p. 8.

[18]. Holeman, I., Cookson, T.P. and Pagliari, C. (2016), "Digital technology for health sector governance in low and middle income countries: a scoping review", Journal of Global Health, Vol. 6 No. 2, pp. 1-11.

[19]. Huff, C.W., & Barnard, L. (2008). Good computing: Life stories of moral exemplars in

the computing profession. IEEE Computers and Society, 28(3), 47-54.

[20]. Huff, C. (2010). What does knowledge have to do with ethics.

[21]. Jacobs, K. (2017), "The ethics of gathering employee data", HR Magazine, 21 March, available at: https://www.hrmagazine.co.uk/content/features/the-ethics-of-gathering-employee-data

[22]. Jin, J., Gubbi, J., Marusic, S. and Palaniswami, M. (2014), "An information framework for creating a smart city through internet of things", IEEE Internet of Things Journal, Vol. 1 No. 2, pp. 112-121, doi: 10.1109/JIOT.2013.2296516.

[23]. Jones, D., Molitor, D. and Reif, J. (2019), "What do workplace wellness programs do? Evidence from the Illinois workplace wellness study", The Quarterly Journal of Economics, Vol. 134 No. 4, pp. 1747-1791.

[24]. Kim, P. (2017), "Data-driven discrimination at work", William and Mary Law Review, Vol. 48, pp. 857-936.

[25]. Kirke, M. (2019), "AI in HR: the good, the bad and the scary", The People Space, 29 May, available at:

[26]. https://www.thepeoplespace.com/leaders/articles/ai-hr-good-bad-and-scary

[27]. Land, F., & Urooj, A. (2003). The ethics of knowledge management. International Journal of Knowledge Management, 3(1), 1-9.

[28]. Manyika, J. (2019), "Tackling bias in artificial intelligence (and in humans)", McKinsey, 6 June, available at: https://www.mckinsey.com/featured-insights/artificial-intelligence/ tackling-bias-in artificial-intelligence-and-in humans#:~:text=AI%20can%20help%20reduce%20bias,bake%20in%20and%20scale%20bias&text=In%20many% 20cases%2C% 20AI%20can,on%20the%20training%20data%20used.

[29]. Nunn, J. (2018), "How AI is transforming HR departments", Forbes, 9 May, available at: https://www.forbes.com/sites/forbestechcouncil/2018/05/09/how-ai-is-transforming-hr-departments/ (accessed 3 March 2021).

[30]. Thoegersen L. Jennifer and Borlund Pia (2021), "Researcher attitudes toward data sharing in public data repositories: a meta-evaluation of studies on researcher data sharing", Journal of Documentation Vol. 78 No. 7, 2022 pp. 1-17 Emerald Publishing Limited.

[31]. Tursunbayeva Aizhan , Pagliari Claudia and Lauro Di Stefano and Antonelli Gilda (2021), "The ethics of people analytics: risks, opportunities and recommendations" Italy Personnel Review Vol. 51 No. 3, 2022 pp. 900-921.

[32]. Tursunbayeva, A., Bunduchi, R., Franco, M. and Pagliari, C. (2016), "Human resource information systems in health care: a systematic evidence review", Journal of the American Medical Informatics Association, Vol. 24 No. 3, pp. 633-654.

[33]. Vermanen, M. and Harkke, V. (2019), "Findings from multipurpose IoT solution experimentations in Finnish SMEs: common expectations and challenges", Proceedings of the 52nd Hawaii International Conference on System Sciences, pp. 5246-5255, doi: 10.24251/HICSS.2019.631.

[34]. Vermanen Mikko, Rantanen M.Minna and Harkke Ville (2021) "Ethical framework for IoT deployment in SMEs: individual perspective", Internet Research Vol. 32 No. 7, 2022 pp. 185-201 Emerald Publishing Limited.

[35]. Wortmann, F. and Fleuchter, K. (2015), "Internet of things", Business and Information Systems Engineering, Vol. 57 No. 3, pp. 221-224, doi: 10.1007/s12599-015-0383-3.

Hindawi
Advances in Materials Science and Engineering
Volume 2022, Article ID 7488254, 16 pages
https://doi.org/10.1155/2022/7488254

Research Article

Experimental Analysis of Geopolymer Concrete: A Sustainable and Economic Concrete Using the Cost Estimation Model

Manvendra Verma,[1] **Kamal Upreti**,[2] **Prashant Vats**,[3] **Sandeep Singh**,[3] **Prashant Singh**,[4] **Nirendra Dev**,[1] **Durgesh Kumar Mishra**,[5] **and Basant Tiwari**[6]

[1]*Department of Civil Engineering, Delhi Technological University, Delhi, India*
[2]*Department of Computer Science & Engineering, Dr. Akhilesh Das Gupta Institute of Engineering & Technology, New Delhi, India*
[3]*Department of Computer Science and Engineering, Faculty of Engineering and Technology, SGT University, Gurugram, Haryana, India*
[4]*Department of Computer Science & Engineering, JSS Academy of Technical Education, Noida, India*
[5]*Department of Computer Science & Engineering, Sri Aurobindo Institute of Technology, Indore, India*
[6]*Deparent of Computer Science IoT, Hawassa University, Hawassa, Ethiopia*

Correspondence should be addressed to Basant Tiwari; basanttiw@hu.edu.et

Received 30 July 2022; Revised 14 August 2022; Accepted 17 August 2022; Published 6 October 2022

Academic Editor: Samson Jerold Samuel Chelladurai

Copyright © 2022 Manvendra Verma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Geopolymer concrete is sustainable, economical, eco-friendly, durable, and high-strength concrete. Geopolymer is a name for the bonding that occurs during the binding of materials in alkaline conditions. Due to the presence of high silica and alumina content, pozzolanic materials could be used as binding materials in the GPC. This research aims to check the sustainability and cost analysis of both GPC and conventional concrete with their physical, chemical, and mechanical properties. The experimental investigation analyzes both GPC and OPC concrete's physical, chemical, and mechanical properties for the M30 mix design and analyzes the concrete's cost and sustainability. The experimental investigation shows that the setting time, density, and drying shrinkage of conventional concrete are higher than the GPC. The compressive strength of the GPC and OPC concretes both showed similar trends at the 28-day strength, but the initial three-day strength of the GPC concrete was much higher than the OPC concrete. The splitting tensile strength and flexural strength of the GPC specimens are slightly higher than the OPC concrete mix specimens. The OPC concrete's elastic modulus is slightly higher than the GPC mix design, whereas the Poisson's ratio of the OPC concrete is slightly lower than the GPC specimens. The GPC specimens have higher thermal stability up to 800°C. The GPC utilizes industrial solid waste like fly ash and slag as a binding material and is activated by an alkaline solution containing NaOH and Na_2SiO_3 in the design mix. Therefore, the GPC has less embodied energy compared to the OPC concrete. The cost of the GPC at a bulk level reduced the cost of up to 40% of the OPC concrete.

1. Introduction

GPC has become a perfect alternative to the world's sustainable construction industry because the concrete demand is the second largest in the world after water. The expansion of infrastructure accelerates the development of the country and society. So, concrete production will increase exponentially in the future with development. A comparable amount of carbon is emitted into the environment during the production of OPC cement. In the GPC, industrial solid waste, fly ash, and slag are used. Fly ash is generated in thermal power plants as a waste product in vast amounts, but the utilization of fly ash is limited to up to 50% of the production [1]. GPC shows better physical, chemical, and mechanical properties compared to OPC cement concrete. The durability properties of the GPC are also higher than those of OPC concrete. GPC concrete is more sulfate-resistant than OPC concrete [2–5]. The geopolymerisation

reaction plays a vital role in developing the strength of the GPC, and some factors affect the geopolymerisation process, like curing conditions, alkaline solution content, and binding material content in the design mix.

Due to the high silica and alumina content composition, pozzolanic materials like fly ash, GGBFS, metakaolin, and rice husk ash were used as binding materials in the geopolymer. The compressive strength increases with the Si/Al ratio [6]. The sodium hydroxide and sodium silicate mix solutions were used to activate the pozzolanic material to work as a binding material for the geopolymerisation reaction. The GPC mix's strength increases with NaOH's molarity increment, reducing workability [7–12]. The geopolymerisation reaction develops the concrete's strength, which is affected by various factors and parameters. The alkaline ratio also affects the mix's strength, increasing with the Na_2SiO_3/NaOH ratio [13].

The curing condition is an essential parameter for gaining strength, and the strength increases with the increment of the curing temperature up to 100°C [14, 15]. The replacement of fly ash by the GGBFS develops the GPC mix's strength in ambient curing conditions [16]. The fineness of the flux plays a vital role in the development of strength; it increases strength and decreases the mix's setting time [17, 18]. The fineness of fly ash reduces porosity directly by increasing the number of small particles in the mix [19, 20]. In the geopolymerisation reaction, curing time affects the mix's early strength development, and a longer curing time gives better end products after the reaction [21]. The water present in the mix was not used in the geopolymerisation end products, so water evaporation made the specimens crack-free [22, 23]. As a result, the water content of the mix increases, reducing the strength of the mixed specimens, and slag has a high potential for reactivity and bond formation for strength development [9, 24]. The GPC has a very high potential to resist the acidic or sea-water environment and other extreme weather conditions [25], and it shows the future scope of the geopolymer paste as a repair of the concrete structure due to its efficiency and performance [26]. GPC specimens are more resistant to high temperatures than OPC concrete samples [27]. Manufacturing sand or stone dust has a high capacity for fine aggregate in concrete without lowering its strength. It also contributes to the formation of high-strength GPC [28, 29]. All alternative materials used in building construction components have a lower energy level than conventional building materials [28–46].

Embodied energy is used to calculate the carbon footprints of various material production methods [47]. The GGBFS and fly ash content in the concrete's mix design reduce the concrete's embodied energy [48, 49]. The optimum point for mechanical properties is a GGBFS/Flash ratio of 20/80 [46]. The SNF-based superplasticizer is efficient for the mechanical properties of GPC [37], and the liquid-to-binder ratio plays a vital role in strength development [34]. The PCE-based superplasticizer is used to self-compact conventional concrete [31]. The study successfully reused waste rubber tyres and waste wood ash by developing sustainable geopolymer concrete that could be utilized for construction industry applications to create various construction and concrete products [50]. GPC is more resistant to various durability conditions than OPC concrete [33, 51, 52].

2. Experimental Program

2.1. Binding Materials. All the constituents used in forming the GPC and OPC concrete specimens for the experimental analysis are explained separately in the paragraphs. The preliminary test on the material samples was to identify the concrete laboratory properties in the civil engineering department, DTU, Delhi. The SEM and EDS tests of the samples were done in the nanotechnology lab, Jamia Millia Islamia, New Delhi, whereas the XRD test was done in the central facility laboratory, DTU, Delhi. The OPC 43 grade of cement was purchased from JK Cement to experiment with a controlled mix design [53]. In the preliminary check, the cement's quality is determined by testing consistency, initial and final setting time, specific gravity, fineness of the particle size, and soundness. The cement passes all preliminary tests as per Indian standard codes [56–62]. The cement's properties are identified through the testing of samples in the concrete laboratory in the civil engineering department, DTU, Delhi. Fly ash is an industrial solid waste produced by a thermal power plant from coal ash fumes and electrostatic precipitation. The particle size of the fly ash is slightly smaller or similar to that of the OPC particle size, and the composition contains a high concentration of silica and alumina. Fly ash was brought from the national thermal power plant, Dadri, Gautam Budh Nagar, Uttar Pradesh, India.

GGBFS is produced in the steel manufacturing plant by quenching the slag from the molten iron steel material. After separating the waste from steel manufacturing, the waste material present in the iron or steel ores is slag. GGBFS also contains a high silica and alumina content in the composition. GGBFS was brought from the Bhilai steel plant in Bhilai, Chhattisgarh, India, to test and produce the GPC mix specimens. Figure 1(a) shows the XRD graph of the GGBFS, which shows the sample's amorphous nature, whereas Table 1 shows the chemical composition present in the GGBFS sample. In addition, Figure 1(b) describes the SEM image of the GGBFS sample at a resolution of two microns and shows the irregular shape of the particles, whereas Figure 1(e) shows the EDS graph, which describes the elements present in the samples.

The fly ash used in the test is a mix-design class-c fly ash [54]. The particles of the fly ash are spherical and porous, as confirmed by the SEM image. Figure 1(a) shows the amorphous nature of the fly ash sample in the XRD graph, whereas Figure 1(c) shows the porous spherical particles of the fly ash in the SEM image at a 5-micron resolution. Table 1 shows the composition of the chemical constituents present in the fly ash as explained by the XRF test, and Figure 1(d) shows the EDS graph, which describes the element content present in the fly ash as conducted in the nanotechnology lab at Jamia Millia Islamia, New Delhi, India.

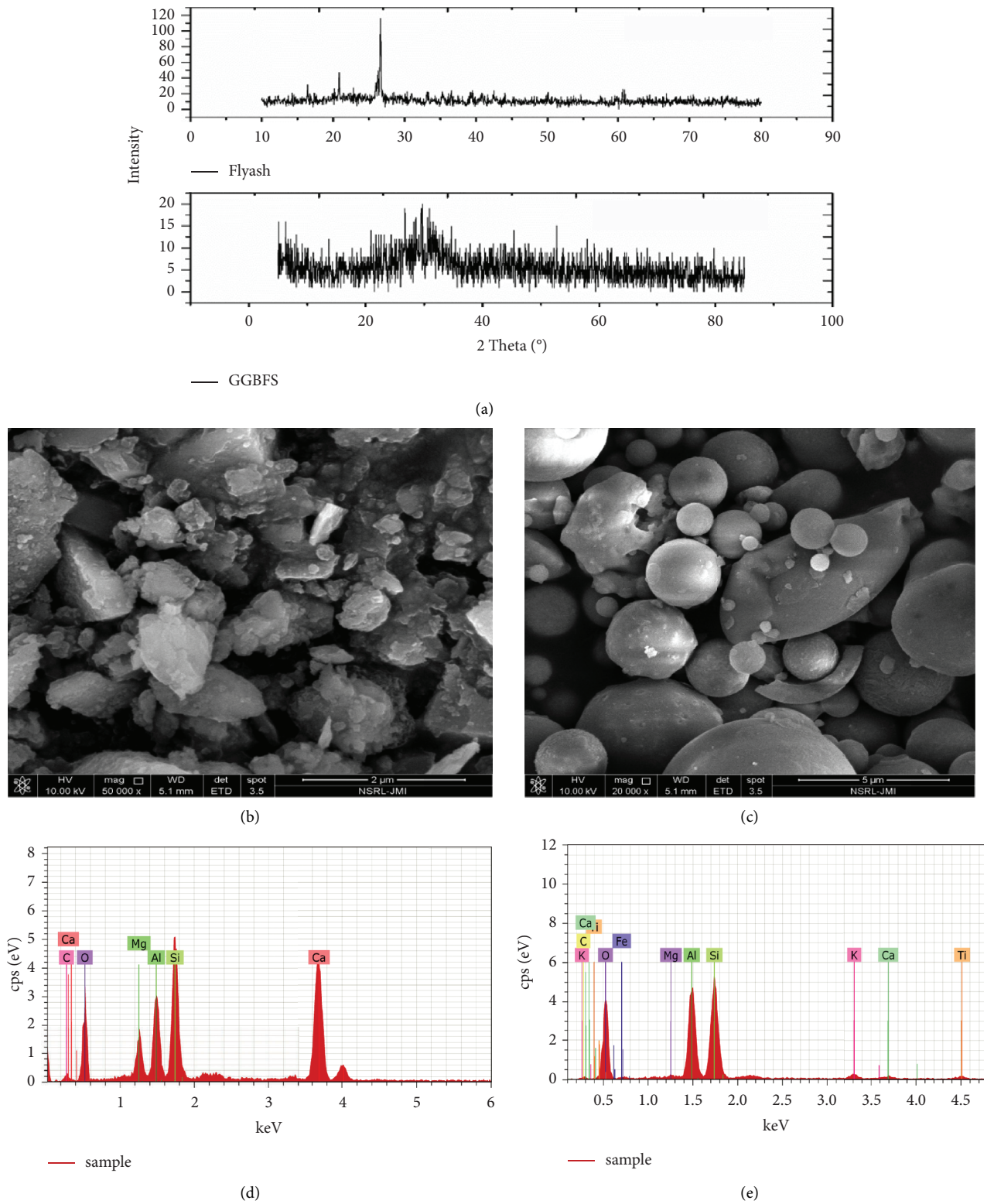


Figure 1: (a) XRD pattern of fly ash and GGBFS; (b) SEM image of GGBFS; (c) SEM image of fly ash; (d) EDS graph of fly ash; and (e) EDS graph of GGBFS.

2.2. Alkaline Solution. The alkaline solution plays a vital role in the geopolymerisation reaction because it activates the reaction's pozzolanic binding materials and finds the end products. For sampling, the alkaline solution contains sodium hydroxide and sodium silicate solution, and it is mixed for 20–24 hours. Sodium hydroxide was purchased from Fisher Scientific in the business park, Powai, Mumbai, Maharashtra, India. Figure 2(a) shows the sample of sodium

Table 1: Composition of fly ash and GGBFS.

Characteristics	SiO_2	Al_2O_3	CaO	Fe_2O_3	MgO	SO_3	LOI
Flyash (%)	45.8	21.4	13.7	12.6	1.3	1.9	.1
GGBFS (%)	34.52	20.66	32.43	.57	10.09	.77	.3

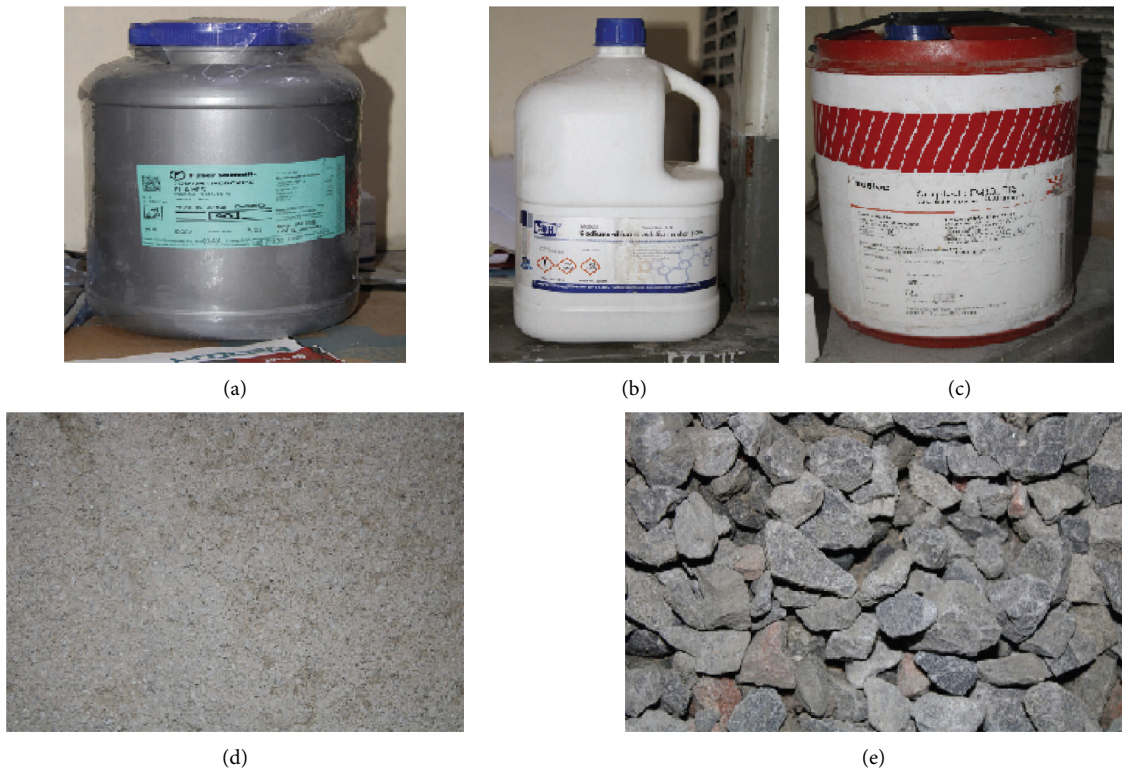


Figure 2: Picture of raw materials of the GPC (a) sodium hydroxide flakes, (b) sodium silicate solutions, (c) superplasticizer, (d) stone-dusts, and (e) coarse aggregates.

hydroxide flakes used in the mix design, whereas Figure 2(b) shows the sodium silicate sample purchased from the Central Drug House (P) Ltd.

2.3. Aggregates. Aggregate is used as the concrete's skeleton because it occupies up to 85% of the concrete. The aggregate used in the mix is primarily classified into two categories: fine aggregates and coarse aggregates. In the coarse aggregate, the two types of aggregates used are 10 mm and 20 mm in the design mix, whereas the fine aggregate is used as crushed stone dust in the mix of both concrete. To test the aggregate use quality in the mix design using Indian standard codes. In the preliminary test, check the gradation of the aggregates, zone, fineness modulus, specific gravity, water absorption, silt content, bulk density, crushing value, impact value, abrasion value, flakiness index, and elongation index of the aggregate samples [55–60]. Figure 2(d) depicts the raw material of stone dust used as a fine aggregate in the mixed design. Figure 3 describes the grain-size curve on the logarithmic graph and the particle size distribution or gradation of the m-sand by the sieve analysis. The stone-dust properties were found through the tests conducted in the concrete lab.

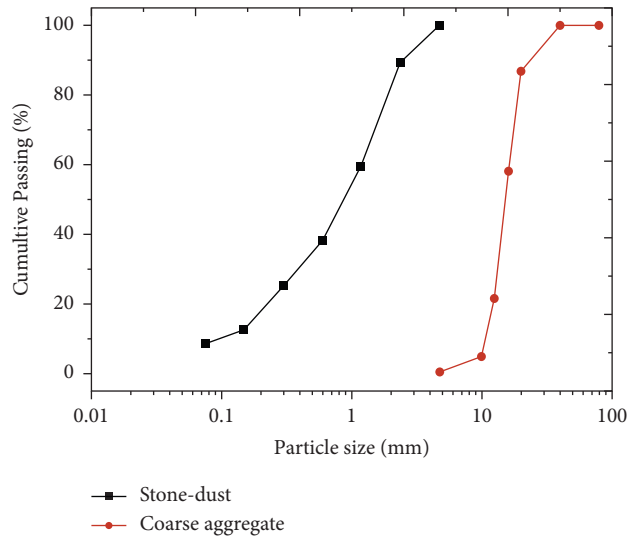


Figure 3: Grain size distribution curve.

As the percentage passing the 600μ sieve is between 35 and 59, the sand belongs to gradation II. From the gradation curve, we find $D_{10} = 0.1$, $D_{30} = 0.4$, and $D_{60} = 1.2$.

Table 2: Mix proportion of GPC and OPC concrete.

Constituents	OPC concrete mix content (kg/m^3)	Geopolymer concrete mix content (kg/m^3)
OPC	370	00
Flyash	00	303.75
GGBFS	00	101.25
NaOH	00	40.5
Na_2SiO_3	00	101.25
Fine aggregate	683	683
Coarse aggregate	1289	1269
Water	148	40.5
Superplasticiser	3.7	4.05
Total	2493.7	2543.7

$$\text{Cu} = D_{60}/D_{10} = 1.2/0.1 = 12 > 6.$$

$$C_c = (D_{30})^2/(D_{10} \text{ x } D_{60}) = 0.4^2/(0.1 \times 1.2) = 1.33.$$

Thus, the sand is well-graded.

Locally available materials, coarse aggregates, are used in all mixes of concrete. The particle size of aggregates present in the sample was determined by their percentage and found the coarse aggregate's fineness modulus. The coarse aggregate properties were discovered through laboratory tests under Indian standards. Figure 2(e) depicts coarse aggregate sample images, whereas coarse aggregate properties are discovered through laboratory tests following Indian standards. The fineness modulus of the coarse aggregate is 7.29, calculated through the sieve analysis of the samples.

2.4. Superplasticiser. The superplasticiser enhances the concrete's performance by reducing the mix design's water content and increasing the fresh mix's workability [61]. The SNF-based superplasticizer is used in the mix design of both concrete made by the Fosroc industry named SP Conplast-430 in the market. Figure 2(c) shows that the picture of the superplasticizer used in the experimental investigation.

2.5. Sampling and Test Setups. Specimens made of the controlled OPC concrete mix usually mix all constituents into the pan mixture for around 2 to 5 minutes and are cast in the specimen mould for 24 hours. In GPC, the alkaline solution mixes before 20–24 hours of mixing the concrete constituents in the pan mixture and is cast in the mould of the specimens with the proper compaction of the casted samples [62–65]. The OPC concrete specimens were cured in the water tank, but the GPC specimens were cured in the oven at 60°C for 24 hours. The cubical, cylindrical, and beam-shaped specimens were cast out of both types of concrete for testing. Physical and mechanical properties were tested for both types of concrete. Table 2 describes the mixed proportion of all constituents used in GPC and OPC concrete's mixed design.

2.6. Physical Properties. The slump and compaction factor tests are used for the identification of the workability of the concrete. Slump is most commonly used for testing the workability of concrete on-site or in the laboratory, whereas the compaction factor is only used in the laboratory. The slump is conical in shape, but the compaction factor setup comprises two conical-shaped buckets fitted in vertical alignment with a standard gap, allowing the buckets' surface to be opened. The compaction factor checks the self-compaction ability of the concrete with gravity force [64, 66]. Self-compacting concrete is made using a PCE-based superplasticizer in the concrete mix design.

2.6.1. Chemical Properties. The weight of the specimens was calculated from the density of both mixed specimens before the destructive tests. Thus, the cube specimen's weight determines the mixed density 28 days after the casting, and the mass calculates the density-to-volume ratio of the cube specimens [67].

2.6.2. Mechanical Properties Tests. The compressive strength of both types of concrete mixes was tested by the cube sample test under the CTM machine at the 5.25 kN/sec rate of the loading statically applied to the specimens. The cube sizes of 150 mm × 150 mm x 150 mm follow the Indian standard code. The mixed samples were tested 3, 7, 14, and 28 days after the specimens were cast [68, 69].

Splitting tensile is used to find the indirect tensile strength of the concrete. According to the Indian standard codes used to determine the splitting tensile of concrete mix specimens, the cylindrical shape of the size diameter x length is 150 mm × 300 mm. For the splitting tensile test of the concrete, a loading rate of 4.5 kN/sec was applied in the transverse direction of the cylindrical specimens [70]. As a result, the splitting tensile is greater than the direct tensile but less than the flexural strength of the same concrete mix design.

Flexural strength, also known as concrete rupture, is used to determine the bending capability of concrete specimens. If the maximum aggregate size is less than 20 mm, a beam of the standard size of 100 mm × 100 mm x 500 mm is used to analyze the flexural strength of the cast specimens. A two-point load was applied along the specimens' transverse direction for the test on the flexural testing machine [71].

The elastic modulus and the Poisson's ratio of the concrete mix are analyzed by testing on the cylindrical specimens. First, the uniaxial static load was applied vertically to the cylindrical specimens and found the cylinder's vertical and horizontal displacement and strength. Then, the

Poisson's ratio was calculated through the horizontal strain ratio of the vertical strain from the displacements. The elastic modulus is found through the load applied to the cylindrical specimens at about one-third of their strength and release, and going continuously through the same procedure often draws the stress-strain graph and finds the elastic modulus through the chord modulus as per the American standard code. Figure 4 shows the mechanical properties test pics of 4(a) compressive strength, 4(b) splitting tensile, 4(c) flexural strength, and 4(d) Poisson's ratio and elastic modulus.

2.7. Nondestructive Tests. The nondestructive test is also used to identify the specimen's strength without any destruction, and it primarily uses the rebound hammer test and the UPV test. The rebound hammer test based on the specimen's surface penetration is reflected by the specimen's surface hardness [72]. The UPV test propagates the ultrasonic pulse in the sample and measures the time travel duration to pass the specimen. The transducers are connected on two opposite sides of the specimen, where one works as an emitter and the other works as a receiver of ultrasonic pulse waves as per the Indian standard [73]. Figure 5 shows the nondestructive test pics of (a) rebound strength and (b) UPVT.

3. Results and Discussion

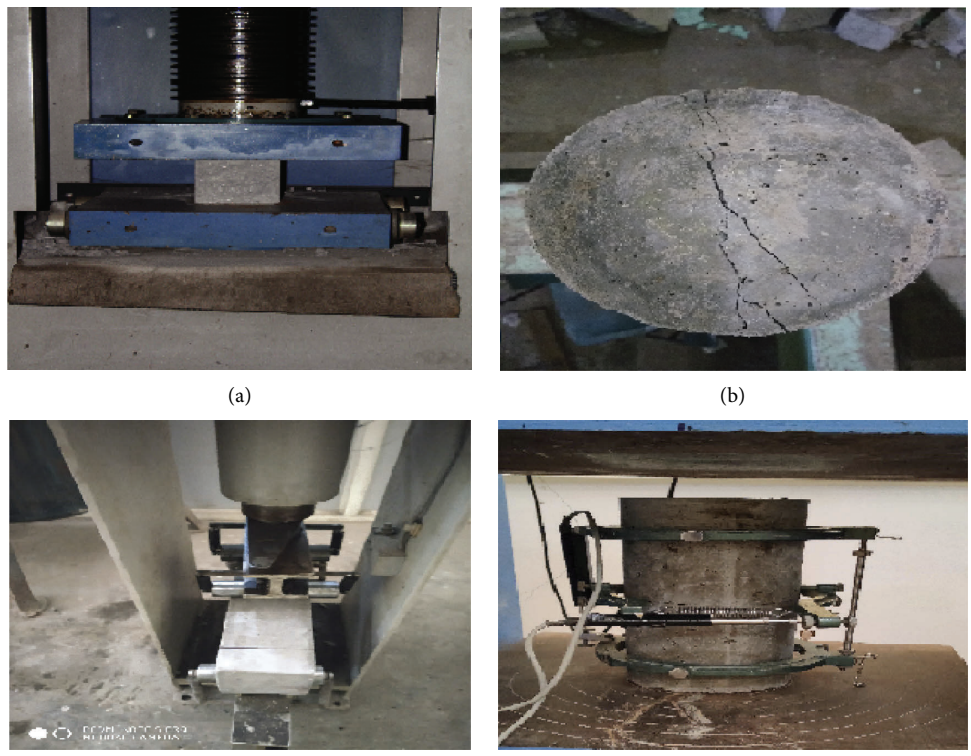
All the cast specimens of both concrete mixes were tested in the machines compared to the concrete mix's mechanical properties. In addition, the specimens of cubical, cylindrical, and beams were tested as per the Indian standards. The experimental work on GPC and OPC concrete mixes was conducted in the concrete laboratory of the Civil Engineering Department, Delhi Technological University, Delhi. It includes the slump value, compaction factor, setting time, density, drying shrinkage, compressive strength, splitting tensile, flexural strength, Poisson's ratio, elastic modulus, rebound strength, and UPV.

3.1. Physical Properties. The slump and compaction factor tests are used to analyze the workability of the fresh concrete mix. In both types of concrete, the slump values of both OPC and GPC concrete are the same at 75–100 mm, and the compaction factor of the OPC concrete is .89, whereas the GPC is .87 after mixing in the pan mixture for 3–5 minutes. The concrete mix's workability is increased by adding a superplasticizer to the mix without reducing the concrete's strength. The SNF-based superplasticizer does not harm the GPC specimen's strength [74]. It also increases microstructure development [75], but it is not beneficial at elevated temperatures [85–89]. The workability of the GPC decreases with the increment of the fineness of the fly ash [76], whereas it increases with the increment of the ratio of alkaline solution to fly ash [77]. The workability decreases with an increment in the alkaline ratio or molarity of the sodium hydroxide [78]. The setting time of the GPC is higher than that of conventional concrete. Figure 6(a) shows both the initial and final setting times. The setting time of the GPC decreases with the increment in the molarity of sodium hydroxide and GGBFS content in the mix design [79, 80].

3.2. Chemical Properties. The concrete sample cube's weight is used to calculate density before the compressive strength test after 28 days of casting. The density of the OPC concrete mix design increases with time, whereas the density of the GPC decreases at the same time. The maximum density of GPC is 2492 kg/m^3 found three days after the casting, whereas the maximum density of OPC concrete is 2462 kg/m^3 found 28 days after the casting. The concrete density with age is the inverse case in both concretes because GPC density decreases with increasing time while OPC density increases. Figure 6(c) shows the graph between the density and time in which the GPC sample density decreases with increasing time, but the OPC concrete density increases after up to 28 days of testing. The density of the concrete is increased by the addition of finer content in the mix-like nanoparticle materials. The drying shrinkage of conventional concrete is higher than that of GPC. Figure 6(b) shows the graph of the drying shrinkage of both concrete specimens. The conventional concrete specimens' drying shrinkage is initially lower than GPC, but it increases and lowers the GPC specimens' drying shrinkage.

3.3. Compressive Strength. The compressive strength of GPC and OPC concretes is similar in trends at 28 days' strength, but the GPC strength is much higher in the first three days' strength than the OPC concrete, but the compressive strength of the concrete mix is similar around 28 days. The three-day compressive strength of GPC and OPC concrete is 23.2 MPa and 14.4 MPa, respectively, while the 28-day compressive strength of GPC and OPC concrete is 35 MPa and 35.6 MPa. Figure 6(d) depicts the compressive time graph for both concretes. The graph shows that the initial strength of the GPC is higher than the OPC concrete samples but shows similar strength at 28 days of compressive strength. The compressive strength of OPC concrete is increased by reducing water content, the addition of superplasticizer, an increment in cement content, and silica fumes in the mix design. The compressive strength of the GPC depends on various parameters such as the molarity of NaOH, alkaline ratio, curing condition, curing temperature, curing time, water content, and superplasticizer. The compressive strength of the GPC increases with the increment of NaOH concentration in the mix design [81, 82]. It increases with the increment of the GGBFS content, curing temperature, curing period, mixing period, and fineness of fly ash [18, 83].

3.4. Splitting Tensile. The splitting tensile strength of the concrete mix designs was found by testing the mix design's cylindrical specimens in the compression testing machine at a loading rate of 4.5 kN/sec statically. The splitting tensile strength of the GPC is higher compared to the OPC concrete mix specimens. At 3, 7, 14, and 28 days, the splitting tensile of the OPC concrete is 1.5 MPa, 3 MPa, 3.5 MPa, and



(a) (b)

(c) (d)

Figure 4: Mechanical properties test pictures of (a) compressive strength; (b) splitting tensile; (c) flexural strength; and (d) Poisson's ratio and elastic modulus.



(a) (b)

Figure 5: Nondestructive tests pics of (a) rebound strength and (b) UPVT.

3.8 MPa, respectively, whereas the splitting tensile of the GPC is 2.6 MPa, 3.2 MPa, 3.9 MPa, and 4.2 MPa, respectively. Thus, the splitting strength of the GPC is slightly higher than that of the OPC concrete for the same compressive strength mix. Figure 6(e) depicts the time-dependent splitting tensile of both concretes. It shows that the GPC samples have higher splitting tensile compared to the OPC concrete samples. The GPC specimens show an initial high strength after 7 days, but after 28 days, they show similar trends. As the sand/fly ash ratio is increased, the tensile strength gradually decreases [84]. The comparison of ACI standard code design strength and experimental test results of geopolymer concrete tensile strength revealed that the splitting tensile strength of GPC matched the design

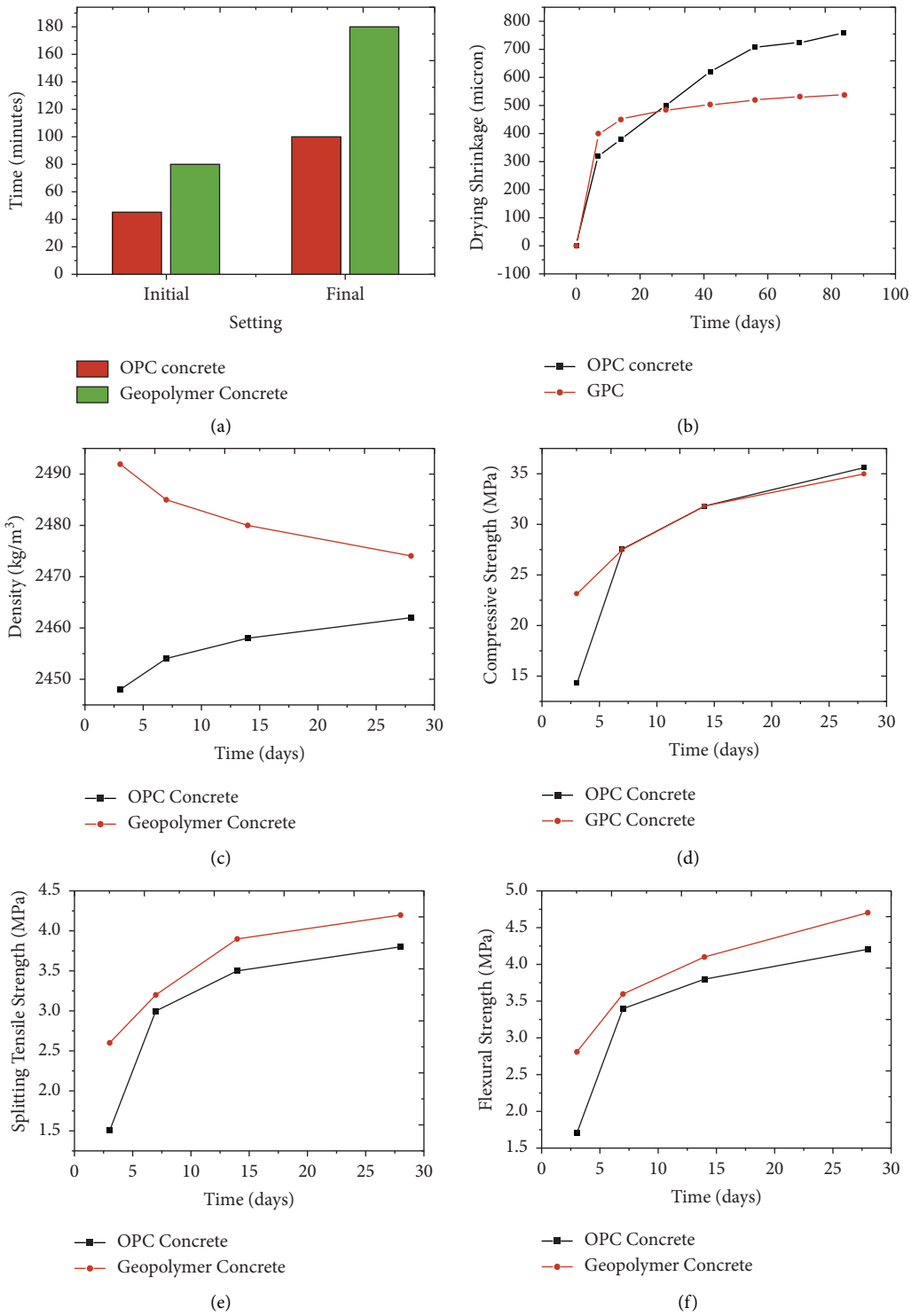


Figure 6: (a) Setting time of concretes; (b) drying shrinkage of the concretes; (c) density variations with time; (d) compressive strength of GPC and OPC concrete; (e) splitting tensile of the GPC and OPC concrete; and (f) flexural strength of GPC and OPC concrete.

specifications specified by the ACI standard [85]. In terms of tensile strength, the geopolymer concrete outperformed the OPC concrete [86, 87].

3.5. Flexural Strength. The flexural strength of the concrete mix designs is identified by testing the beam samples of the mix design in the flexural testing machine under a two-point load applied on the section of the beam in the transverse direction of the specimens. The flexural strength of the GPC specimens is slightly higher than the OPC concrete mix specimens. The flexural strength of OPC concrete is 1.7 MPa, 3.4 MPa, 3.8 MPa, and 4.2 MPa after 3, 7, 14, and 28 days, respectively, whereas the GPC flexural strength is 2.8 MPa,

3.6 MPa, 4.1 MPa, and 4.7 MPa after 3, 7, 14, and 28 days. Figure 6(f) describes the flexural strength and the time of both concretes. It shows that the GPC has higher flexural strength than OPC concrete samples in all tests. With normal curing, the inclusion of 5–6% nano-silica enhanced the tensile strength of GPC [88]. The GPC experimental findings using GGBFS and fly ash revealed that using GGBFS in the GPC matrix had a good effect on the direct tensile strength of GPC [89]. According to the literature research, the tensile strength of GPC concrete is superior to that of OPC concrete with the same compressive strength [90].

3.6. Elastic Modulus and Poisson's Ratio. The elastic modulus and the Poisson's ratio of the concrete mix design are identified by testing the cylindrical specimens of the mix in the compression testing machine with vertical and horizontal extension meters. The load was applied along the vertical direction of the specimens statically in the machine. To calculate the lateral and linear deformation of specimens subjected to strain after failure, the lateral strain ratio to the linear strain shows the Poisson's ratio of the concrete mix design, whereas the load applied one-third of the failure load to the specimens, reduced the load, and continued the same procedure many times. The slope of the initial stress-strain curve describes the elastic modulus of the concrete mix design. The GPC and OPC concrete mix designs' elastic modulus was tested 28 days after casting the specimen moulds. The OPC concrete's elastic modulus is slightly higher than the GPC mix design, whereas the Poisson's ratio of the OPC concrete is slightly lower than the GPC mix design. The elastic modulus of the OPC concrete and GPC is 26.3 GPa and 23.4 GPa, respectively, whereas the Poisson's ratio of the OPC concrete and GPC is .15 and .17, respectively. Figure 4(d) shows the setup to determine the Poisson's ratio and elastic modulus of the concrete mix cylindrical samples.

3.7. Nondestructive Tests

3.7.1. Rebound Strength. The rebound strength test is generally used in the field as a nondestructive test. It is based on the hammer striking on the specimen's surface with some IS code provisions and shows the rebound value. The rebound hammer is struck on the specimen surface ten times continuously, and the rebound gives the average rebound value and the rebound strength of the specimens according to Indian standards. The rebound strength shows similar trends to the mix's compressive strength, but the rebound strength shows a little higher value. For example, the initial three-day strength of GPC is higher than the OPC concrete mix but shows somewhat lower strength than the OPC concrete. Figure 7(a) describes the graph of the rebound strength of both GPC and OPC concrete.

3.7.2. UPV Test. The UPV test is vital in the nondestructive test method because it provides strength and crack information for the existing structures. The UPV test needs only two opposite sides of specimens to propagate ultrasonic pulse waves in a clean condition. Because rebound tests are based on surface hardness, they are only used on specimens with a thickness of up to 300 mm. The UPV test is primarily used in consulting work on existing concrete structures. The UPV of the specimens increases with hardness, which is directly proportional to their strength. Figure 7(b) describes the graph of the UPV test results of both GPC and OPC concrete mix samples. It shows similar trends in GPC and OPC concrete to the rebound strength graph.

3.8. TGA (Thermo Gravimetric Analysis). The TGA test does the thermal analysis of materials by increasing the temperature by 10°C per second up to 900°C and checking the derivative of weight reduction with the temperature. The apparatus shows notes of the weight reduction of the material with increasing temperature in a graphical form. The two parameters are represented in a single graph with different y-axes but a single x-axis, where one y-axis represents the derivative weight percentage and the other represents the weight percentage reduction with temperature. The GPC matrix is highly stable at elevated temperatures without failure compared to the OPC concrete matrix. Figure 8 describes the graph of TGA-DTG of GPC concrete up to the temperature of 850°C. It shows that the weight decreases with increasing temperature, but the GPC matrix is highly stable up to 850°C, and it shows the retained material is 92% of the original at 850°C. The DTG graph shows that the weight reduction with temperature is nonlinear, varying from negative to positive.

4. Mechanical Properties Correlation

The correlation is created by analyzing data on mechanical properties obtained from destructive laboratory tests. The linear regression analysis of compressive strength, splitting tensile, and flexural tensile strength showed the correlation equation and compared it to the Indian standard correlation equations.

4.1. Correlation between the Compressive Strength and Splitting Tensile. The correlation between splitting tensile and compressive strength is proposed in equation (1). It is produced as a result of regression analysis of data generated by destructive testing in the laboratory. In the state-of-the-art of high-strength concrete [91], the ACI committee gives equation (2), and a committee of the European Code International gives equation (3) in the model code of 1990 [92]. Equation (4) was introduced by an ACI standard and report for building code requirements for structural concrete [93]. The graph of the correlation between compressive strength and splitting tensile strength is shown in Figure 9. A regression analysis of the test data yields the correlation equation.

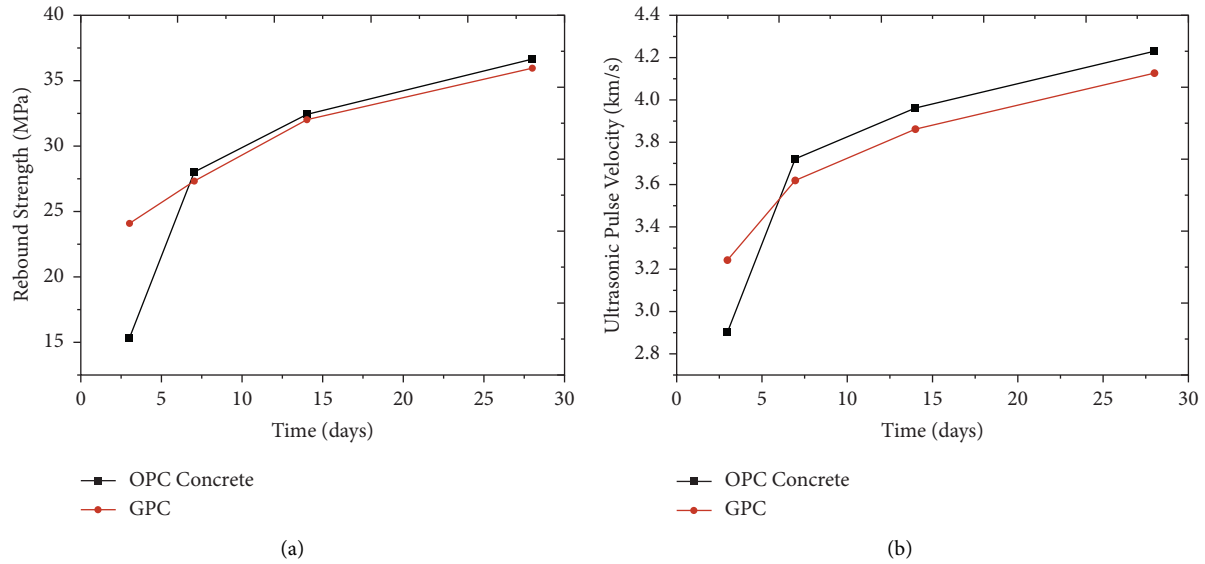


Figure 7: (a) Graph of OPC concrete and GPC rebound strength; and (b) graph of OPC concrete and GPC UPV test.

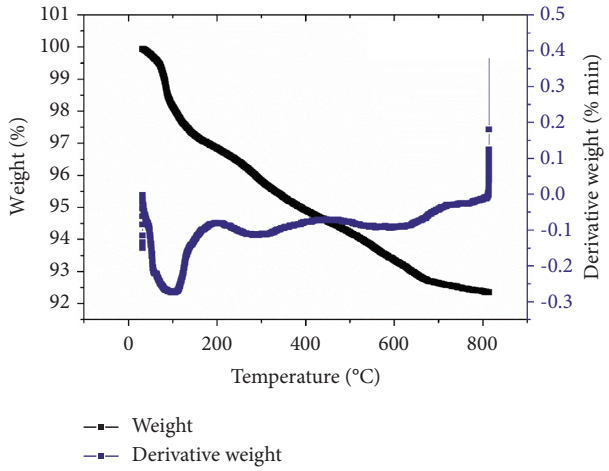


Figure 8: Graph of TGA-DTG test of GPC mix.

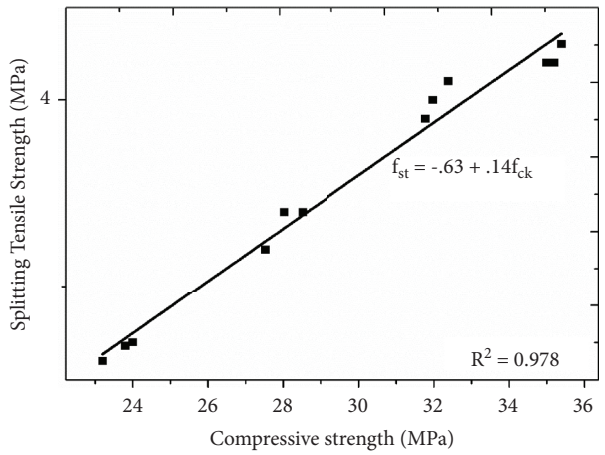


Figure 9: Graph of correlation between compressive strength and splitting tensile.

$$\text{Propose d: } f_{st} = -63 + 14 f_{ck}, \tag{1}$$

$$ACI363R - 92: \; f_{st} = 0.59\sqrt{f_c}, \tag{2}$$

$$CEB - FIP: \; f_{st} = 0.301 f_c^{0.67}, \tag{3}$$

$$ACI318 - 14: \; f_{st} = 0.56 f_c^{0.5}. \tag{4}$$

4.2. Correlation between Compressive Strength and Flexural Strength. The correlation between the compressive strength and flexural description by equation (5) is through the regression analysis of the test data. In the publication of the state-of-the-art of high strength concrete, the ACI committee included equation (6). A standard report presented on the code for the design of concrete in a structure introduces the equation (7) of correlation. The Australian code for reinforced concrete design presented equation (8), and the Indian code introduced equation (9) between flexural strength and compressive strength [94, 95]. The flexural strength is always about 10%–20% of the compressive strength. The tested data also lie between these values. The graph of the correlation equation between flexural strength and compressive strength proposed by the tested data is shown in Figure 10.

$$\text{Propose d: } f_{fs} = -1.03 + 16 f_{cs}, \tag{5}$$

$$ACI363R - 92: \; f_{fs} = 0.94\sqrt{f_c}, \tag{6}$$

$$ACI318 - 99: \; f_{fs} = 0.62\sqrt{f_c}, \tag{7}$$

$$AS3600: \; f_{fs} = 0.6\sqrt{f_c}, \tag{8}$$

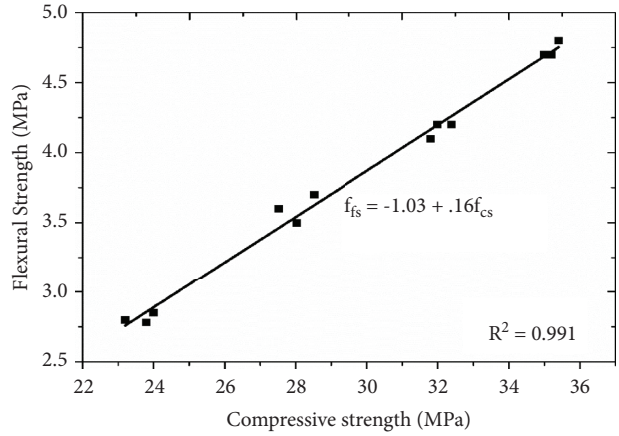


Figure 10: Graph of correlation between compressive strength and flexural strength.

Table 3: Embodied energy calculation of GPC and OPC.

Constituents	Embodied energy (MJ/kg)	OPC concrete		Geopolymer concrete	
		Mix content (kg/m^3)	Embodied energy content (MJ/kg)	Mix content (kg/m^3)	Embodied energy content (MJ/kg)
OPC	4.6	370	1702	—	—
Fly ash	0.0	—	—	303.75	00
GGBFS	0.2	—	—	101.25	31.38
NaOH	20.5	—	—	40.5	830.25
Na_2SiO_3	5.37	—	—	101.25	543.71
Fine aggregate	0.02	683	13.66	683	13.66
Coarse aggregate	0.22	1289	283.58	1269	279.18
Water	0.0	148	00	40.5	00
Superplasticiser	12.6	3.7	46.62	4.05	51.03
	Total	2493.7	2045.86 MJ/m^3	2543.7	1749.21 MJ/m^3

$$IS456 - 2000: \ f_{fs} = 0.7\sqrt{f_c}. \tag{9}$$

5. Sustainability Analysis

Sustainable development is essential in the present scenario. This vast volume of industrial solid waste is now created, necessitating this trash in development activities. The use of industrial solid waste decreases carbon footprints and, as a result, indirectly lowers the use of high embodied energy items. For example, the embodied energy in the dry process of OPC cement is approximately 4.6 MJ/kg, whereas fly ash and slag have no embodied energy. By replacing the OPC, industrial solid waste fly ash and slag are used as the binder in concrete, and those wastes are activated by a chemical solution of sodium hydroxide and sodium silicate. As a result, the GPC has less embodied energy than the OPC concrete [96–98]. Table 3 shows the embodied energy calculation for both concrete and concrete. The embodied energy of all constituents of concrete is calculated based on the reference papers.

6. Cost Analysis

The concrete cost was calculated by the number of the constituents present in the concrete mix design, and then, after calculating the rate of all individual constituent prices, the concrete mix design cost was found. The cost of the GPC at a bulk level reduced the cost of up to 40% of the OPC concrete. The cost of the OPC of 1 m^3 is Rs. 3758, whereas the cost of the GPC of 1 m^3 is Rs. 2230. The cost estimate only includes the materials and excludes the human resources. Table 4 describes all the calculations of the cost of the materials and concrete. Materials and concrete costs are critical in the construction industry because construction projects are always more expensive. The project's economy always matters and goes to the optimum point, which is essential for the project's lower cost. When compared to OPC concrete, GPC concrete is less expensive and emits less carbon dioxide. This is a new trend in the construction industry due to sustainable development properties. Thaarrini and Dhivya compared the cost of GPC to cement concrete.

OPC concrete is 11% more expensive than GPC for better quality concrete. GPC costs 1.7% more than cement concrete for grades up to 30 MPa [99]. Through his

Table 4: Cost analysis of the GPC and OPC concrete.

Constituents	Cost rate (Rs./kg)	OPC concrete		Geopolymer concrete	
		Mix content (kg/m^3)	Cost of the constituent ($Rs./m^3$)	Mix content (kg/m^3)	Cost of the constituent ($Rs./m^3$)
OPC	7.6	370	2812	—	—
Fly ash	0.7	—	—	303.75	212.62
GGBFS	2.5	—	—	101.25	253.13
NaOH	10.05	—	—	40.5	407.03
Na_2SiO_3	10	—	—	101.25	1012.5
Fine aggregate	0.25	683	170.75	683	170.75
Coarse aggregate	0.525	1289	676.72	1269	666.23
Water	0.0	148	00	40.5	00
Superplasticiser	26.67	3.7	98.68	4.05	108.01
	Total	2493.7	Rs. $3758.15/m^3$	2543.7	Rs. $2230.27/m^3$

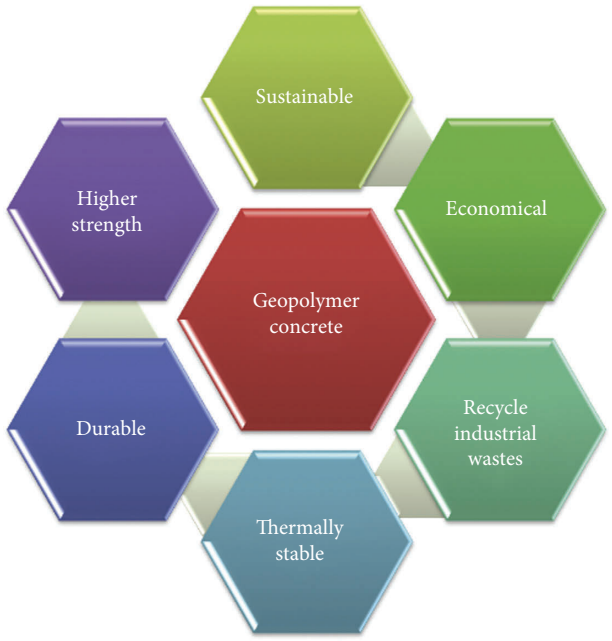


Figure 11: Picture of short description of the GPC conclusion.

experimental analysis, McLellan et al. discovered that using fly ash-based GPC decreases building costs and greenhouse gas emissions by 7% and 64%, respectively. The cost of producing GGBS GPC was discovered to be 7% more than the cost of producing OPC-based concrete [100, 101]. Assi et al. developed a novel mix design approach for fly ash-based GPC that reduced construction costs by 50% [102].

7. Conclusions

After the experimental, sustainability, and cost investigations, they concluded the following:

(i) In both types of concrete, the slump value of both OPC concrete and GPC is the same, 75–100 mm, and the compaction factor of the OPC concrete is .89, whereas the GPC is .87 after mixing in the pan mixture for 3–5 minutes. The setting time of the GPC specimens is higher than the OPC concrete specimens.

(ii) The density of the OPC concrete mix design increases with time, whereas the density of the GPC decreases at the same time, and the drying shrinkage of the OPC concrete specimens is higher than the GPC specimens, but initially, its lower value before the 28-days test.

(iii) The compressive strength of the GPC and OPC concretes is similar in trends at the 28-day strength but not at the initial three-day strength; the GPC strength is much higher than the OPC concrete, but the compressive strength of the concrete mix becomes more similar around 28 days.

(iv) The splitting tensile and flexural strength of the GPC is slightly higher than the OPC concrete mix specimens. Similarly, the OPC concrete's elastic modulus is slightly higher than the GPC mix design, whereas the Poisson's ratio of the OPC concrete is slightly lower than the GPC mix design.

(v) The rebound strength shows similar trends to the mix's compressive strength, but the rebound strength shows a little higher value. The UPV results show similar trends in GPC and OPC concrete to the rebound strength graph.

(vi) The TGA test shows that the weight decreases with increasing temperature, but the GPC matrix is highly stable up to 850°C, and it shows the retained material is 92% of the original at 850°C.

(vii) The GPC has less embodied energy compared to the OPC concrete.

(viii) The cost of the GPC at a bulk level reduced the cost of up to 40% of the OPC concrete. The cost of the OPC concrete of 1 m^3 is Rs. 3758, whereas the cost of the GPC of 1 m^3 is Rs. 2230.

(ix) The experimental results proposed that the correlation equation between compressive strength and splitting tensile was $f_{st} = -.63 + .14f_{ck,}$ and the correlation equation between compressive strength and flexural strength was $f_{fs} = -1.03 + .16f_{cs.}$

(x) After the research, the conclusion is represented in the graph form in Figure 11.

Data Availability

All data are included in the article.

Additional Points

The authors also focus on a brief investigation of geopolymer concrete in an application. The microstructural characterization of these hybrid reinforced composites with low calcium geopolymer concrete is performed.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The experimental testing was supported by the Civil Engineering Department, Delhi Technological University, New Delhi, India.

References

[1] Central Electricity Authority, *Flyash Generation at Coal/ lignite Based thermal Power Stations and its Utilization in the Country*, 2019.

[2] I. Ismail, S. A. Bernal, J. L. Provis, S. Hamdan, and J. S. J. van Deventer, "Microstructural changes in alkali-activated fly ash/slag geopolymers with sulfate exposure," *Materials and Structures*, vol. 46, no. 3, pp. 361–373, 2013.

[3] C. Chotetanorm, P. Chindaprasirt, V. Sata, S. Rukzon, and A. Sathonsaowaphak, "High-calcium bottom ash geopolymer: sorptivity, pore size, and resistance to sodium sulfate attack," *Journal of Materials in Civil Engineering*, vol. 25, no. 1, pp. 105–111, 2013.

[4] W. G. Valencia Saavedra, D. E. Angulo, and R. Mejía de Gutiérrez, "Fly ash slag geopolymer concrete: resistance to sodium and magnesium sulfate attack," *Journal of Materials in Civil Engineering*, vol. 28, no. 12, pp. 1–9, 2016.

[5] M. A. R. Bhutta, W. M. Hussin, M. Azreen, and M. M. Tahir, "Sulphate resistance of geopolymer concrete prepared from blended waste fuel ash," *Journal of Materials in Civil Engineering*, vol. 26, no. 11, pp. 1–6, 2014.

[6] P. Duxson, J. L. Provis, G. C. Lukey, S. W. Mallicoat, W. M. Kriven, and J. S. van Deventer, "Understanding the relationship between geopolymer composition, microstructure and mechanical properties," *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, vol. 269, no. 1-3, pp. 47–58, 2005.

[7] G. Nagalia, Y. Park, A. Abolmaali, and P. Aswath, "Compressive strength and microstructural properties of fly ash – based geopolymer concrete," *Journal of Materials in Civil Engineering*, vol. 28, no. 12, pp. 1–11, 2016.

[8] G. S. Ryu, Y. B. Lee, K. T. Koh, and Y. S. Chung, "The mechanical properties of fly ash-based geopolymer concrete with alkaline activators," *Construction and Building Materials*, vol. 47, pp. 409–418, 2013.

[9] J. G. Jang, N. K. Lee, and H. K. Lee, "Fresh and hardened properties of alkali-activated fly ash/slag pastes with superplasticizers," *Construction and Building Materials*, vol. 50, pp. 169–176, 2014.

[10] P. Topark-Ngarm, P. Chindaprasirt, and V. Sata, "Setting time, strength, and bond of high-calcium fly ash geopolymer concrete," *Journal of Materials in Civil Engineering*, vol. 27, pp. 1–7, 2015.

[11] D. Hardjito, S. E. Wallah, D. M. J. Sumajouw, and B. V. Rangan, "On the development of fly ash-based geopolymer concrete," *ACI Materials Journal*, vol. 101, pp. 467–472, 2004.

[12] S. V. Patankar, Y. M. Ghugal, and S. S. Jamkar, "Effect of concentration of sodium hydroxide and degree of heat curing on fly ash-based geopolymer mortar," *Indian Journal of Materials Science*, vol. 2014, pp. 1–6, 2014.

[13] P. Nath and P. K. Sarker, "Effect of GGBFS on setting, workability and early strength properties of fly ash geopolymer concrete cured in ambient condition," *Construction and Building Materials*, vol. 66, pp. 163–171, 2014.

[14] P. Rovnaník, "Effect of curing temperature on the development of hard structure of metakaolin-based geopolymer," *Construction and Building Materials*, vol. 24, no. 7, pp. 1176–1183, 2010.

[15] P. Nath and P. K. Sarker, "Flexural strength and elastic modulus of ambient-cured blended low-calcium fly ash geopolymer concrete," *Construction and Building Materials*, vol. 130, pp. 22–31, 2017.

[16] S. Kumar, R. Kumar, and S. P. Mehrotra, "Influence of granulated blast furnace slag on the reaction, structure and properties of flyash-based geopolymer," *Journal of Materials Science*, vol. 45, no. 3, pp. 607–615, 2010.

[17] A. Sathonsaowaphak, P. Chindaprasirt, and K. Pimraksa, "Workability and strength of lignite bottom ash geopolymer mortar," *Journal of Hazardous Materials*, vol. 168, no. 1, pp. 44–50, 2009.

[18] S. S. Jamkar and Y. M. Ghugal Svp, "Effect of fineness of fly ash on flow and compressive strength of geopolymer concrete," *Indian Concrete journal*, vol. 87, no. 4, pp. 57–62, 2013.

[19] S. Luhar and U. Khandelwal, "A study on water absorption and sorptivity of geopolymer concrete," *International Journal of Civil Engineering*, vol. 2, no. 8, pp. 1–9, 2015.

[20] N. A. Lloyd and B. V. Rangan, "Geopolymer concrete with fly ash," in *Proceedings of the Second International Conference on Sustainable Construction Materials and Technologies*, pp. 1–7, 2010.

[21] D. S. Perera, E. R. Vance, O. Uchida, and K. S. Finnie, "Influence of curing schedule on the integrity of geopolymers," *Journal of Materials Science*, vol. 42, no. 9, pp. 3099–3106, 2007.

[22] D. Bondar, C. J. Lynsdale, N. B. Milestone, and N. Hassani, "Sulfate resistance of alkali activated pozzolans," *International Journal of Concrete Structures and Materials*, vol. 9, no. 2, pp. 145–158, 2015.

[23] D. Bondar, C. J. Lynsdale, N. B. Milestone, and N. Hassani, "Engineering properties of alkali activated natural pozzolan concrete," *Second Int Conf Sustain Constr Mater Technol*, vol. 108, pp. 1–10, 2010.

[24] N. K. Lee, J. G. Jang, and H. K. Lee, "Shrinkage characteristics of alkali-activated fly ash/slag paste and mortar at early ages," *Cement and Concrete Composites*, vol. 53, pp. 239–248, 2014.

[25] K. U. Ambikakumari Sanalkumar, M. Lahoti, and E. H. Yang, "Investigating the potential reactivity of fly ash for geopolymerization," *Construction and Building Materials*, vol. 225, pp. 283–291, 2019.

[26] Y. C. Ding, T. W. Cheng, and Y. S. Dai, "Application of geopolymer paste for concrete repair," *Structural Concrete*, vol. 18, no. 4, pp. 561–570, 2017.

[27] Z. Pan, J. G. Sanjayan, and F. Collins, "Effect of transient creep on compressive strength of geopolymer concrete for elevated temperature exposure," *Cement and Concrete Research*, vol. 56, pp. 182–189, 2014.

[28] K. Upreti, M. Verma, M. Agrawal et al., "Prediction of mechanical strength by using an artificial neural network and random forest algorithm," *Journal of Nanomaterials*, vol. 2022, pp. 1–12, 2022.

[29] A. Chouksey, M. Verma, N. Dev, I. Rahman, and K. Upreti, "An investigation on the effect of curing conditions on the mechanical and microstructural properties of the geopolymer concrete," *Materials Research Express*, vol. 9, no. 5, Article ID 55003, 2022.

[30] R. Kumar, M. Verma, and N. Dev, "Investigation of fresh, mechanical, and impact resistance properties of rubberized concrete," in *Proceedings of the International E-Conference on Sustainable Development & Recent Trends in Civil Engineering*, pp. 88–94, delhi, India, January 2022.

[31] M. Verma and M. Nigam, "Mechanical behaviour of self compacting and self curing concrete," *International Journal Innovation Research Science Engineer Technology*, vol. 6, pp. 14361–14366, 2017.

[32] M. Verma, "Experimental investigation on the properties of Geopolymer concrete after replacement of river sand with the M-sand," in *Proceedings of the International E-Conference on Sustainable Development & Recent Trends in Civil Engineering*, pp. 46–54, delhi, India, January 2022.

[33] R. Kumar, M. Verma, and N. Dev, "Investigation on the effect of seawater condition, sulphate attack, acid attack, freezet-haw condition, and wetting–drying on the geopolymer concrete," *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, vol. 46, no. 4, pp. 2823–2853, 2021.

[34] M. Verma and N. Dev, "Effect of liquid to binder ratio and curing temperature on the engineering properties of the geopolymer concrete," *Silicon*, vol. 14, no. 4, pp. 1743–1757, 2022.

[35] A. Chouksey, M. Verma, N. Dev, I. Rahman, and K. Upreti, "An investigation on the effect of curing conditions on the mechanical and microstructural properties of the geopolymer concrete," *Materials Research Express*, vol. 9, no. 5, Article ID 55003, 2022.

[36] M. Verma and N. Dev, "Geopolymer concrete: a way of sustainable construction," *International Journal of Innovative Research in Advanced Engineering*, vol. 5, pp. 201–205, 2018.

[37] M. Verma and N. Dev, "Effect of SNF-based superplasticizer on physical, mechanical and thermal properties of the geopolymer concrete," *Silicon*, vol. 14, no. 3, pp. 965–975, 2021.

[38] M. Verma and N. Dev, "Geopolymer Concrete: A Sustainable and Economic Concrete via Experimental Analysis," *Research square*, pp. 1–39, 2021.

[39] N. Juneja and K. Upreti, "An Introduction to Few Soft Computing Techniques to Predict Software Quality," in *Proceedings of the 2017 2nd Int Conf Telecommun Networks*, Noida, India, August 2017.

[40] J. Ananthi, N. Sengottaiyan, S. Anbukaruppusamy, K. Upreti, and A. K. Dubey, "Forest fire prediction using IoT and deep learning," *International Journal of Advanced Technology and Engineering Exploration*, vol. 9, pp. 246–256, 2022.

[41] K. Upreti, U. K. Singh, R. Jain, K. Kaur, and A. K. Sharma, "Fuzzy logic based support vector regression (SVR) model for software cost estimation using machine learning," *ICT Systems and Sustainability*, vol. 321, pp. 917–927, 2022.

[42] M. Verma and N. Dev, "Sodium hydroxide effect on the mechanical properties of flyash-slag based geopolymer concrete," *Structural Concrete*, vol. 22, no. S1, pp. E368–E379, 2021.

[43] M. Verma, N. Dev, I. Rahman, M. Nigam, M. Ahmed, and J. Mallick, "Geopolymer concrete: a material for sustainable development in Indian construction industries," *Crystals*, vol. 12, no. 4, p. 514, 2022.

[44] M. Verma and N. Dev, "Review on the effect of different parameters on behavior of Geopolymer Concrete," *Int J Innov Res Sci Eng Technol*, vol. 6, pp. 11276–11281, 2017.

[45] M. Verma and N. Dev, "Effect of superplasticiser on physical, chemical and mechanical properties of the geopolymer concrete," *Challenges of Resilient and Sustainable Infrastructure Development in Emerging Economies*, pp. 1185–1191, Kolkata, India, 2020.

[46] M. Verma and N. Dev, "Effect of ground granulated blast furnace slag and fly ash ratio and the curing conditions on the mechanical properties of geopolymer concrete," *Structural Concrete*, vol. 23, no. 4, pp. 2015–2029, 2021.

[47] G. P. Hammond and C. I. Jones, "Embodied energy and carbon in construction materials," *Proceedings of the Institution of Civil Engineers - Energy*, vol. 161, no. 2, pp. 87–98, 2008.

[48] K. Kupwade-Patil and E. N. Allouche, "Examination of chloride-induced corrosion in reinforced geopolymer concretes," *Journal of Materials in Civil Engineering*, vol. 25, no. 10, pp. 1465–1476, 2013.

[49] K. Kupwade-Patil and E. N. Allouche, "Impact of alkali silica reaction on fly ash-based geopolymer concrete," *Journal of Materials in Civil Engineering*, vol. 25, no. 1, pp. 131–139, 2013.

[50] K. Arunkumar, M. Muthukannan, A. Suresh kumar, and A. Chithambar Ganesh, "Mitigation of waste rubber tire and waste wood ash by the production of rubberized low calcium waste wood ash based geopolymer concrete and influence of waste rubber fibre in setting properties and mechanical behavior," *Environmental Research*, vol. 194, Article ID 110661, 2021.

[51] R. Kumar, M. Verma, N. Dev, and N. Lamba, "Influence of chloride and sulfate solution on the long-term durability of modified rubberized concrete," *Journal of Applied Polymer Science*, pp. 1–15, 2022.

[52] M. Verma, A. Juneja, and D. Saini, "Effect of waste tyre rubber in the concrete," in *International E-Conference on Sustainable Development & Recent Trends in Civil Engineering*, pp. 99–103, 2022.

[53] *IS 8112 1989 (1990) 43 GRADE ORDINARY PORTLAND CEMENT - SPECIFICATION*, Bur Indian Stand, 1990.

[54] ASTM C 618 19, *Standard Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use in concrete*, 2019.

[55] "IS 2386 (Part II) (1998) Methods of test for aggregates for concrete Part II Estimation of deleterious materials and organic impurties," *Bur Indian Stand*, vol. 2386, 1998.

[56] "IS 2386 (Part I) (1997) Methods of test for aggregates for concrete Part I Particle size and shape," *Bur Indian Stand*, vol. 2386, 1997.

[57] "IS 2386 (Part III) Methods of test for aggregates for concrete Part III Specific gravity, density, voids, absorption and bulking," *Bur Indian Stand*, vol. 2386, 1997.

[58] "IS 383 1970 Specification for coarse and fine aggregates from natural sources for concrete," *Bur Indian Stand*, pp. 1–20, 1997.
[59] "IS 2386 (Part IV) Methods of test for aggregates for concrete Part IV Mechanical Properties," *Bur Indian Stand*, vol. 2386, 1997.
[60] IS 2386 (PartV), *Methods of Test for Aggregates for concrete Part V Soundness*, Bur Indian Stand, 1997.
[61] *IS 9103 1999 CONCRETE ADMIXTURES-Specification*, Bur Indian Stand April, 1999.
[62] *IS 12119 1987 GENERAL REQUIREMENT FOR PAN MIXTURES FOR CONCRETE*, Bur Indian Stand, 1999.
[63] *IS 10262 2009 CONCRETE MIX PROPORTIONING - GUIDELINES*, Bur Indian Stand, 2009.
[64] "IS: 1199 - 1959 reaffirmed, IS 1199 1959," *METHODS OF SAMPLING AND ANALYSIS OF CONCRETE*, Bur Indian Stand, 1959.
[65] "IS 7325 1974 (1999) specification for apparatus for determining constituents of fresh concrete," *Bur Indian Stand*, pp. 1–12, 1999.
[66] *IS 7320 1974 Specification for concrete Slump Test Apparatus*, Bur Indian Stand, 2008, https://doi.org/10.1136/archdischild-2011-300172.
[67] "(unit weight), yield, and air content (gravimetric) of concrete," 2019, https://doi.org/10.1520/C0138 ASTM C138/C138M 17a (2019) Standard Test Method for Density ASTM Int 1–6.
[68] *IS 516 1959 METHODS OF TEST FOR STRENGTH OF CONCRETE*, Bur Indian Stand, 2004.
[69] "IS 10086 1982," *SPECIFICATION FOR MOULDS FOR USE IN TESTS OF CEMENT AND CONCRETE*, Bur Indian Stand, 2008.
[70] *IS 5816 1999 SPLITTING TENSILE STRENGTH OF CONCRETE - METHOD OF TEST*, Bur Indian Stand, 1999.
[71] "IS 9399 1979," *SPECIFICATION FOR APPARATUS FOR FLEXURAL TESTING OF CONCRETE*, Bur Indian Stand, 1979.
[72] *IS 13311-2 Method of Non-destructive Testing of concrete-methods of Test Part 2: Rebound Hammer*, Bur Indian Stand, 1992.
[73] IS 13311 (Part1), *Non-destructive Testing of concrete Methods of Test Part 1 Ultrasonic Pulse Velocity*, Bur Indian Stand, 1992.
[74] M. Albitar, P. Visintin, M. S. Mohamed Ali, and M. Drechsler, "Assessing behaviour of fresh and hardened geopolymer concrete mixed with class-F fly ash," *KSCE Journal of Civil Engineering*, vol. 19, no. 5, pp. 1445–1455, 2015.
[75] S. Demie, M. F. Nuruddin, and N. Shafiq, "Effects of microstructure characteristics of interfacial transition zone on the compressive strength of self-compacting geopolymer concrete," *Construction and Building Materials*, vol. 41, pp. 91–98, 2013.
[76] T. Phoo-ngernkham, C. Phiangphimai, N. Damrongwiriyanupap, S. Hanjitsuwan, J. Thumrongvut, and P. Chindaprasirt, "A mix design procedure for alkali-activated high-calcium fly ash concrete cured at ambient temperature," *Advances in Materials Science and Engineering*, vol. 2018, Article ID 2460403, 15 pages, 2018.
[77] C. A. Jeyasehar and M. Salahuddin, "Development of fly ash based geopolymer concrete precast elements," *Ministry of Environment and Forests of India and Annamalai University*, pp. 1–77, 2013.
[78] I. Perná, T. Hanzlicek, and M. Supova, "The identification of geopolymer affinity in specific cases of clay materials," *Applied Clay Science*, vol. 102, pp. 213–219, 2014.
[79] V. D. Cao, S. Pilehvar, C. Salas-bringas et al., "Micro-encapsulated phase change materials for enhancing the thermal performance of Portland cement concrete and geopolymer concrete for passive building applications," *Energy Conversion and Management*, vol. 133, pp. 56–66, 2017.
[80] B. S. Umniati, P. Risdanareni, F. Tarmizi, and Z. Zein, "Workability enhancement of geopolymer concrete through the use of retarder workability enhancement of geopolymer concrete through the use of retarder," in *Proceedings of the AIP Conference Proceedings*, pp. 1–9, Sepetember 2017.
[81] E. Najafi Kani and A. Allahverdi, "Effects of curing time and temperature on strength development of inorganic polymeric binder based on natural pozzolan," *Journal of Materials Science*, vol. 44, no. 12, pp. 3088–3097, 2009.
[82] C. Sreenivasulu, A. Ramakrishnaiah, and J. G. Jawahar, "M echanical P roperties of G eopolymer C oncrete," *International Journal of Advances in Engineering & Technology*, vol. 8, pp. 83–91, 2015.
[83] G. Nagalia, Y. Park, A. Abolmaali, and P. Aswath, "Compressive strength and microstructural properties of fly ash – based geopolymer concrete," *Journal of Materials in Civil Engineering*, vol. 28, no. 12, pp. 1–11, 2016.
[84] W. K. W. Lee and J. D. Deventer, "Chemical interactions between siliceous aggregates and low-Ca alkali-activated cements," *Cement and Concrete Research*, vol. 37, no. 6, pp. 844–855, 2007.
[85] X. Y. Zhuang, L. Chen, S. Komarneni et al., "Fly ash-based geopolymer: clean production, properties and applications," *Journal of Cleaner Production*, vol. 125, pp. 253–267, 2016.
[86] H. Y. Zhang, V. Kodur, B. Wu, L. Cao, and F. Wang, "Thermal behavior and mechanical properties of geopolymer mortar after exposure to elevated temperatures," *Construction and Building Materials*, vol. 109, pp. 17–24, 2016.
[87] Z. H. Zhang, H. J. Zhu, C. H. Zhou, and H. Wang, "Geopolymer from kaolin in China: an overview," *Applied Clay Science*, vol. 119, pp. 31–41, 2016.
[88] D. Adak, M. Sarkar, and S. Mandal, "Effect of nano-silica on strength and durability of fly ash based geopolymer mortar," *Construction and Building Materials*, vol. 70, pp. 453–459, 2014.
[89] A. Hassan, M. Arif, and M. Shariq, "Use of geopolymer concrete for a cleaner and sustainable environment – a review of mechanical properties and microstructure," *Journal of Cleaner Production*, vol. 223, pp. 704–728, 2019.
[90] F. Collins and J. G. Sanjayan, "Strength and shrinkage properties of alkali-activated slag concrete placed into a large column," *Cement and Concrete Research*, vol. 29, no. 5, pp. 659–666, 1999.
[91] H. G. Russell, A. R. Anderson, and J. O. Banning, *State-of-the-Art Report on High-Strength Concrete Reported by ACI Committee 363*, 1997.
[92] C. International, *ceb-fip-model-code-1990-design-code*, 1990.
[93] A. C. I. Committee, *Building Code Requirements for Structural Concrete*, pp. 318–414, 2014.
[94] C. Concrete, *Reinforced Concrete Design in Accordance with AS 3600—2009*, 2009.
[95] *IS 456 2000 PLAIN AND REINFORCED CONCRETE - CODE OF PRACTICE*, Bur Indian Stand, 2000.
[96] B. N. Varsha and S. J. K. R. Saranya, "Embodied energy of aggregates and masonry units produced around Bengaluru,

India," *International Journal of Advanced Science and Engineering Technology*, vol. 6, pp. 42–45, 2018.

[97] M. A. Gonzalez Stumpf, M. P. Kulakowski, L. G. Breitenbach, and F. Kirch, "A case study about embodied energy in concrete and structural masonry buildings," *Revista de la construcción*, vol. 13, no. 2, pp. 9–14, 2014.

[98] S. R. Anvekar, L. R. Manjunatha, S. R. Anvekar, S. Sagari, and K. Archana, "An economic and embodied energy comparison of geo-polymer, blended cement and traditional concretes," *J Civ Eng Technol Res*, vol. 1, pp. 33–40, 2014.

[99] T. Janardhanan, J. Thaarrini, S. Dhivya &, and S. Dhivya, "Comparative study on the production cost of geopolymer and conventional concretes," *International Journal of Civil Engineering Research*, vol. 7, pp. 117–124, 2016.

[100] B. C. McLellan, R. P. Williams, J. Lay, A. van Riessen, and G. D. Corder, "Costs and carbon emissions for geopolymer pastes in comparison to ordinary portland cement," *Journal of Cleaner Production*, vol. 19, no. 9-10, pp. 1080–1090, 2011.

[101] B. J. Mathew, M. Sudhakar, and C. Natarajan, "Strength , economic and sustainability characteristics of coal ash – GGBS based geopolymer concrete," *International Journal of Computational Engineering Research*, vol. 3, pp. 207–212, 2013.

[102] L. Assi, K. Carter, E. E. Deaver, R. Anay, and P. Ziehl, "Sustainable concrete: building a greener future," *Journal of Cleaner Production*, vol. 198, pp. 1641–1651, 2018.

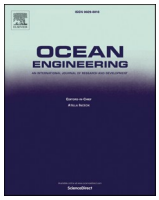
Ocean Engineering 266 (2022) 112838



Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng





Experimental study of the local scour around the two piers in the tandem arrangement using ultrasonic ranging transducers

Geeta Devi *, Munendra Kumar

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

ARTICLE INFO

Keywords:
Bridge pier
Densimetric froude number
GEP
Local scour
Maximum scour depth
Pier spacing

ABSTRACT

Single pier characterization has been predominately considered by a majority of the recommended equations, although all bridge structures are laid on the group of piers. The behaviour of scour must be extensively examined in order to identify the parameters that primarily influence the bridge scour. The primary aim of the study is to investigate the effect of the densimetric Froude number, temporal scour development and pier spacing (P_s) on scour depth and propose a new equation for scour depth estimation using GEP (Gene Expression Programming). For profoundly understanding and continuous monitoring of scour evolution, SeaTek Ultrasonic ranging transducers have been used. The results exemplify that non-dimensional scour depth increases with the increase in the densimetric Froude number and velocity of flow. The effect of pier spacing shows that the scour depth increases as the pier spacing increases at a certain P_s range, then it decreases. The maximum scour depth at the front pier was observed 24% greater than the single pier, and 14% rear pier. The comparison of the proposed GEP equation with the well-known equation and Experimental results shows that the GEP is better prediction tool for scour depth.

1. Introduction

The removal of sediments encircling bridge piers by the action of river flow is known as scour. The development of scour holes surrounding piers and abutments in natural water flow is common phenomenon (Wang et al., 2016a). The scour development encircling the piers can be carried out in two flow conditions: clear-water and live-bed. When the sediment aside from the bridge piers remains unchanged, scour depth on the upstream of the front pier increases as a function of downward flow is known as clear-water scour. When the sediment is moved throughout the entire surface of the bed, since the flow rate is greater than the incipient velocity of sediment, it is known as live-bed scour (Wang et al., 2016a). According to the Jiang and Lin (2022, 2021) and Sumer et al. (2005) scour around the group of piles can be classified as global scour and local scour. Global scour is shallow in depth and wide around the pile group and depression across in the surrounding area, and local scour is steep-sided and depression is only around the individual piles.

Numerous investigations have been performed to determine the factors influencing scour around bridge piers to better understand the complicated behaviour of scour (Bozkuş et al., 2018; Breusers et al., 1977; Devi and Kumar, 2022a; Dey and Raikar, 2007; Hong et al., 2018; Jain et al., 2021; Lança et al., 2013; Ma et al., 2018; Melville and Raudkivi, 1977; Muzzammil and Gangadhariah, 2010; Pandey et al., 2020a; Pu et al., 2021; Tafarojnoruz et al., 2012; Tipireddy et al., 2019; Vijayasree and Eldho, 2021). According to these investigations, scour around bridge piers is among the most significant aspects of the failure of river bridges. The majority of these investigations were centered on single pier scour. Although, in practice, new bridges spans are usually broad, and two or more circular piers are required to support the structure, positioned in the flow direction.

Many new bridges have been built beside older bridges in recent years as the automobile population has increased. Such an example is in Kolkata, West Bengal, India, at which a new bridge, Nivedita Bridge, has successfully opened directly downstream of an existing bridge, Vivekananda Bridge. If two bridges are situated near to one other, the piers of one must have an impact on the piers of the other as shown in Fig. 1. The overall purpose here is to conduct research and experiments adopting tandem pier arrangements of identical pier models, gradually varying the spacing along flow between subsequent circular piers, and then determining significant scouring factors.

As a consequence of rapid economic development in developing countries, the number of bridges constructed has been increased. A

* Corresponding author.
E-mail addresses: geeta_phd2k17@dtu.ac.in (G. Devi), munendrakumar@dtu.ac.in (M. Kumar).

https://doi.org/10.1016/j.oceaneng.2022.112838
Received 10 March 2022; Received in revised form 8 September 2022; Accepted 6 October 2022
Available online 18 October 2022
0029-8018/© 2022 Elsevier Ltd. All rights reserved.

LIST OF SYMBOLS

d	@pier diameter
d_{50}	@mean particle size of the sediment
F_d	@densimetric Froude number
F_r	@Froude number
F_{d84}	@densimetric Froude number at d_{84}
g	@acceleration due to gravity
h	@water depth above the sand bed
ρ	@water density
ρ_s	@density of sediments
P_s	@pier spacing
Q	@rate of flow
R_h	@hydraulic mean depth
σ_g	@geometric standard deviation the particle size
t	@time taken for each experimental evaluation
t_e	@time taken for achieving equilibrium scour depth
u	@averaged velocity of approaching flow
u_c	@critical velocity of the approaching flow
y_s	@scour depth Y_{SU} @scour depth Y_{SU} @scour depth at the upstream side of the pier
Y_{SR}	@scour depth at the right side of the pier
Y_{SL}	@scour depth at the left side of the pier
Y_{SD}	@scour depth at the downstream side of the pier
Y_{t_e}	@percentage of scour depth attained at equilibrium time

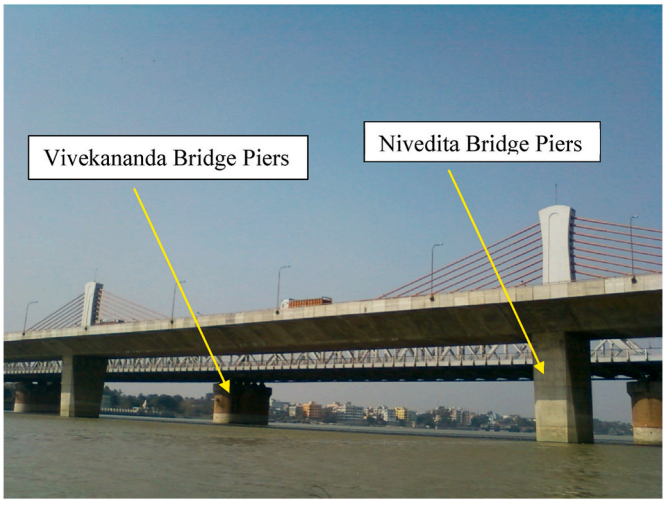


Fig. 1. Two pier in tandem arrangement on the Vivekananda Bridge and Nivedita Bridge.

complete understanding of the scour around pier groups is highly needed. Some of the studies based on two piers in the tandem arrangement were conducted experimentally in order to highlight few important factors for the design consideration of the bridge pier, pier spacing, mutual interference, size of the bridge pier, bridge pier shape and countermeasures such as (Amini et al., 2011, 2014; Beg, n.d.; Beg and Beg, 2015; Behzad and Abolfazl, 2013; Behzad and Beheshti, 2006; Elliott and Baker, 1985; Garg et al., 2021; Hosseini and Amini, 2015; Keshavarzi et al., 2018; Memar et al., 2020) studied the impacts of pier spacing and suggested equations for the side-by-side arrangement to compute scour depth. The study conducted by Behzad and Beheshti (2006) presents a correction factor for estimating the scour depth around the group of piles. By taking into account uncertainties caused by randomness in some parameters, Homaei and Najafzadeh (2020) suggested a methodology to evaluate the correlation between the wave parameters, seabed sediment conditions, pile groups geometry, and the wave-induced-scour depth around pile installed in the marine ecosystem.

Amini et al. (2012) experimentally studied the scour depth under clear water conditions and concluded the dependency of scour on the diameter of pile, spacing between piles, and the ratio of an immersion. Amini et al. (2012) also developed a technique for predicting scour depth for a wide range of an immersed and unimmersed group of piles. In addition, there were few studied such as (Hamidi and Siadatmousavi, 2018; Pandey et al., 2020b; Qi et al., 2019; Suk et al., 2014), are based on numerical analysis for bridge structures, for example, piers, spurs and abutments etc. An experimental investigation was conducted by Ghodsi et al. (2021) to assess the effects of the pile cap on longitudinal extension from the column, the layout of the pile group, the extension that occurs upstream of the pile group, and the thickness of the pile cap. The findings of their investigation indicate that the maximum scour depth reduced, the number of piles in line with flow rose, and variations in pier-foundation design had a major impact on the maximum scour depth.

The influence of the approaching flow characteristics (depth of flow and velocity of flow) and sediment characteristics (density of particle, size of particles, and gradation effect) is subjected to the variation in local scour depth. An increase in non-uniformity of sediment leads the creation of an armored layer at the foot of the scour hole, lowering the scour depth in clear-water scour. The decrease is clearly linked to the gradation factor or geometrical standard deviation, σ_g, which is defined as $\sqrt{\frac{d_{84}}{d_{16}}}$, where d_{84} and d_{16} signify the grain size at the eighty-fourth and sixteenth percentiles, respectively (Cheng et al., 2020). The influence of densimetric Froude number on maximum scour depth and scour deposit has not been explored directly by researchers. It ought to be highlighted that the design equation does not contain characterization of the densimetric Froude number. Additionally, the maximum scour depth also depends on geometric characteristics (pier diameter of piers, spacing between piers, and width of piers) (Zhou et al., 2020). However, until now, researchers have been incapable of finding out the behaviour of the pier, the influence of the pier spacing, influence of densimetric Froude number and temporal scour depth. Table 1 lists previously conducted research work for the two piers, along with their finding and founding.

Predictive approaches based on the artificial Intelligence such as Artificial Neural Networks (ANN) Adaptive Neuro-Fuzzy Inference Systems (ANFIS), Neuro-fuzzy Group Method of Data Handling (Neuro-fuzzy GMDH), Gene Expression Programming (GEP), Evolutionary Polynomial Regression (EPR), Model Tree (MT), and Multivariate Adaptive Regression Splines (MARS) have been studied by Azamathulla (2012); Azamathulla et al. (2008); Azamathulla and Jarrett (2013); Azmathullah et al. (2005); Najafzadeh (2015); Najafzadeh and Azamathulla (2015), to yield effective estimates of scour around hydraulic structures. Azmathullah et al. (2005) used ANN for the estimation of scour depth on the downstream of ski-jump bucket. Azamathulla et al. (2008) used neural networks to estimate the scour below spillways. Azamathulla (2012) has developed a predictive models for the abutment scour using GEP technique and concluded that the GEP models showed satisfactory estimations when compared to with the Artificial Neutral Network-Radial Basis Function (ANN- RBF) modulus. Whereas, Azamathulla and Jarrett (2013) assessed the potential of gene-expression programming (GEP) for estimating the Manning's n values. Najafzadeh (2015) and Najafzadeh and Azamathulla (2015) used Neuro-fuzzy GMDH systems to evaluate the scour depth around the piers. Additionally, Sharafati et al. (2020) utilized GMDH stochastic modelling

Table 1
Research work available in the literature for scour around two bridge piers.

Pier Models	Reference	Investigation	Finding	Founding		
				Y_t	F_d	P_s
One	Cheng et al. (2020)	Assessing the sediment gradation effects under clear water for bridge pier scour	An increase in the grading of sediment particles results in reducing the scour depth.	–	C	–
Two	Wang et al. (2016a)	Time development and flow variation of the scouring depth	Scour Zone of the rear pier was divided into four new categories.	C	–	–
Two	Keshavarzi et al. (2018)	The significance of pier spacing over scour depth was investigated experimentally.	The findings indicate maximum scour depth at the front pier at spacing = 2.5 D.	C	–	C
Two	Behzad and Abolfazl (2013)	This Experimental study compares flow patterns on a rough surface for a single pier and two piers.	The findings show that a rear pier significantly affects the flow pattern, especially in the vicinity zone.	–	–	C
Two	Das et al. (2015)	The study experimentally investigates the scour development and the influence of pier shape and mutual interference on the scour.	The greater scour hole was generated when circular piers were replaced with the equilateral triangular and square pier of the same dimensional width under same flow conditions.	–	–	C
Two and Three	Khaple et al. (2017)	An experimental study, to evaluate the pier for, two piers positioned in tandem and side by side, and three piers positioned in staggered arrangements.	Findings of the study, represents the influence of pier spacing on different arrangements.	C	–	C
Two	Garg et al. (2021)	Experimental investigation of the influence of collar plate on the individual bridge pier and combined collar plate for two piers.	The results of the study, concludes that the collar plate of diameter 2.5 D for single pier and two piers, reduces the scour depth completely.	–	–	C
Two	Memar et al. (2020)	The objective of the experimental study was to evaluate the, the influence of the size and positioning of circular collars over the pier, on scour for two piers in tandem position.	The study concludes that, the location of the collar plate on the bed surface results in maximum scour reduction.	C	–	–
Two	Beg and Beg (2015)	The study investigates factors affecting the scour depth under clear water scour conditions. Including mutual interference between two bridge piers of unequal sizes and pier spacing,	The results conclude that, maximum scour was observed at 66 mm upstream of the front pier and minimum at 33 mm rear piers.	–	–	C
Present Study				C	C	C

Where Y_t is temporal scour depth; F_d is the densimetric Froude number and P_s is the pier spacing factor.
*C conducted.

technique to predict the local scour depth around the single pier in cohesive bed materials.

However, researchers have utilized GEP to model the scour prediction around the single pier. Najafzadeh et al. (2016) and Najafzadeh and Oliveto (2021) developed nonlinear regression equations by utilizing four AI models including GEP, EPR, MT, and MARS to predict the equations for local scour depth around pile groups. Statistical assessment results obtained, shows that, the MARS approach provides more accurate prediction and equations obtained with GEP and MT models revealed convincing efficiency. Hassan and Jalal (2021); Mohammed Saleh et al. (2019); Muzzammil et al. (2015) used GEP to formulate scour depth around single pier. Hassan et al. (2022) formulated expression for predicting the scour depth around the pile group by utilizing GEP, ANN Py-Torch. Whereas (Devi and Kumar, 2022b), used GEP to formulate the scour depth around two pier in the side by side arrangement. Results of these studies shows that the GEP model indicated higher estimation accuracy compared to the established existing regression models.

By seeing the potential of the AI based methodologies for the prediction of the scour depth around the hydraulic structures and the only disadvantage of ANN that, it cannot provide a user friendly simplified equation that is practical for use by pavement engineering practitioners. This is one of the main reasons why ANN methods are considered "black boxes". However, GEP methods provides equations that could be used to develop spreadsheets that are simple for use by practitioners. Through examination of literature, the approach using GEP for the estimation of scour depth for the tandem pier arrangements, has not been used. So, in order to address this problem, present research work was conducted for two piers positioned in the direction of the flow at the various pier spacing. The objectives of this paper are to (a) provide new experimental data for maximum scour depth (MSD) for a single pier and two piers, (b) analyze the effect of pier spacing on MSD, (c) explain the temporal evolution of scour hole, (d) provide the influence of densimetric Froude number on scour depth, and (e) to assess the maximum scour depth and pier spacing effect using gene expression programming (GEP).

2. Dimensional analysis

There is a number of parameters on which scour depth is dependent such as flow characteristics, sediments properties, and bridge pier characteristics. To understand the scour depth, the relationship of these parameters with the scour depth was studied through dimensional analysis, as it was difficult to study analytically. Dimensional analysis is an effective tool for investigating the functional relationship of the

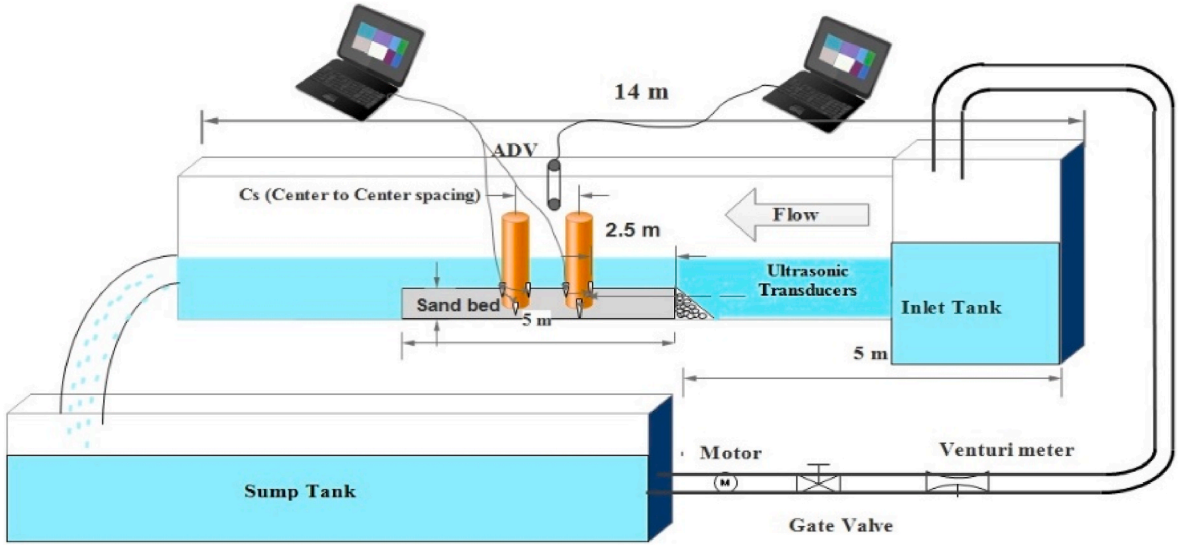


Fig. 2. Experimental setup of hydraulic flume.

influencing parameters. To simplify this analysis, a few assumptions were made, such as: sediment used was uniformly graded, there was no viscous effect, and the effect of approach flow depth was negligible. In the functional relationship for these conditions, the equilibrium scour depth is written as:

$$y_s = f\left(\rho, \rho_s, g, h, d, u, u_c, d_{50}, P_s, t, t_e\right) \tag{1}$$

where, y_s = scour depth;ρ = water density; ρ_s = density of sediments; g = acceleration due to gravity; d = pier diameter; u = averaged velocity of approaching flow; u_c = critical velocity of the approaching flow; d_{50} = mean particle size of the sediment; P_s = *pier spacing*; t = time taken for each experimental evaluation; and t_e = time taken for achieving equilibrium scour depth.

From Eq. (1), dimensionless parameters were obtained using the Buckingham π method of dimensional analysis. Three repeating variables were $h, u, and\ \rho$ representing the geometric, flow and fluid parameters.

$$\Pi_1(g) = \frac{u^2}{gh}, \Pi_2(u_c) = \frac{u}{u_c}, \Pi_3(\rho_s) = \frac{\rho}{\rho_s}, \Pi_4(d_{50}) = \frac{h}{d_{50}}, \Pi_5(d) = \frac{h}{d}, \Pi_6(P_s) = \frac{h}{P_s}\ \ \Pi_7(t) = \frac{h}{t}, \Pi_8(t_e) = \frac{h}{t_e} \tag{2}$$

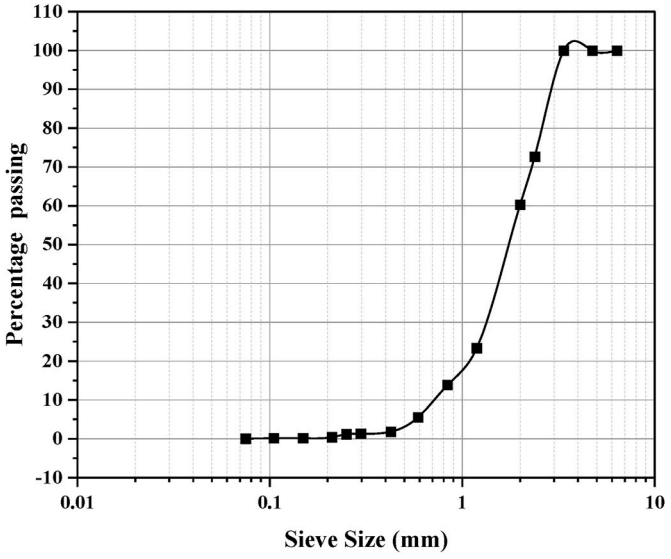


Fig. 3. Particle size distribution curve.

$$\Pi_4(d_{50}) \times \frac{1}{\Pi_5(d)} = \frac{h}{d_{50}} \times \frac{1}{h/d} = \frac{d}{d_{50}},\ \Pi_5(d) \times \frac{1}{\Pi_6(P_s)} = \frac{h}{d} \times \frac{1}{h/P_s} = \frac{P_s}{d},\ \Pi_8(t_e) \times \frac{1}{\Pi_7(t)} = \frac{h}{t_e} \times \frac{1}{h/t} = \frac{t}{t_e} \tag{3}$$

By re-arranging these variables, the non-dimensional relationship can be obtained as:

$$\frac{y_s}{d} = f\left(\frac{u^2}{gh}, \frac{u}{u_c}, \frac{d}{d_{50}}, \frac{P_s}{d}, \frac{t}{t_e}\right) \tag{4}$$

where $\frac{u^2}{gh}$ is defined as Froude number (F_r), which represents the intensity of the bulk flow in comparison to the settling velocity of the particles. $\frac{u}{u_c}$ is the relative velocity; $\frac{t}{t_e}$ is the ratio of the time taken for an experimental duration to the time taken to achieve the equilibrium scour depth;$\frac{P_s}{d}$ is the non-dimensional pier spacing factor; and $\frac{d}{d_{50}}$ is the same for all the flow conditions. Therefore, the impact of three independent dimension-less variables, namely F_d, u/u_c, t/t_e, P_s/d has been studied in this paper.

3. Experimental setup

Experimental work has been carried out in the hydraulic laboratory using recirculating masonry flume at Delhi Technological University, New Delhi, India. Fig. 2 depicts a schematic diagram of the flume detailing the flow sections. The flume was 14 m long, 1.10 m wide, and 0.80 m deep. The flow rate was measured using an Acoustic Doppler Velocimeter (ADV) and a venturimeter. The flume has a settling

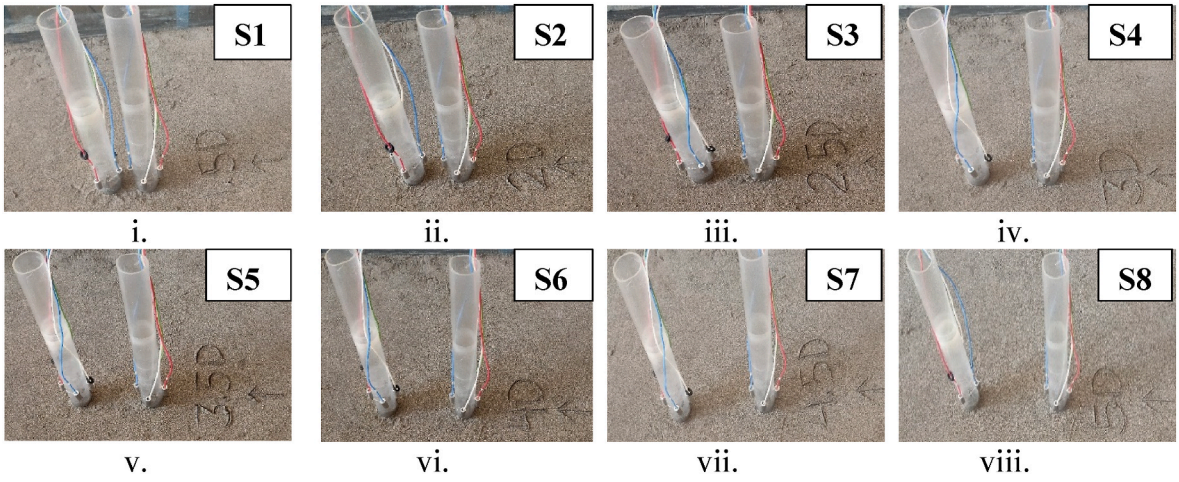


Fig. 4. Pier positioning and observation points.

Table 2
Flow parameters.

Test classification	d (mm)	d_{50} (mm)	h (m)	u (m/sec)	Q (m^3/sec)	F_r	u_c (m/sec)	u/u_c	R_h
A	50	1.75	0.120	0.224	0.0295	0.206	0.235	0.955	2.069
B	50	1.75	0.135	0.243	0.0361	0.211	0.239	0.974	2.109
C	50	1.75	0.149	0.262	0.0428	0.217	0.243	0.987	2.141
D	50	1.75	0.154	0.287	0.0486	0.233	0.244	0.996	2.151
E	50	1.75	0.163	0.300	0.0537	0.237	0.247	0.980	2.168

where, d = pier diameter; d_{50} = mean particle size; h = water depth above the sand bed; u = averaged velocity of approaching flow; Q = rate of flow; F_r = Froude number; u_c = critical velocity of the approaching flow; R_h = hydraulic mean depth.

Table 3
Characteristics flow conditions of the present study and similar studies.

Study	d_{50}	h	σ_g	d	F_{d84}
Ettema (1976)	0.55–4.1	0.6	1.3–4.6	0.1	1.7–3.4
(Albert, 2004)	0.75	0.29–0.40	1.4–3.4	0.18	0.9–3.8
Present study	1.75	0.120–0.163	1.68	0.05	1.4–1.8

where, d_{50} = mean particle size; h = depth of water above the sand bed; σ_g = geometric standard deviation the particle size, d = pier diameter; F_{d84} = densimetric Froude number at d_{84} (84 percentage of the particles are finer than the whole sample).

Table 4
Outline of the dimensional factors utilized in the formulation of the GEP.

Variables	Name	Range
V_1	Diameter of pier (D)	5 cm
V_2	Mean-Particle size of sediment used (d_{50})	1.75 mm
V_3	Flume width	1.10 m
V_4	Working section length	5 m
V_5	Flow Depth (H)	0.12–0.16 m
V_6	Flow Velocity (U)	0.224–0.30 m/s
V_7	Pier Spacing (P_s)	1.5D-5D
V_8	Critical Velocity(U_c)	0.308–0.322 m/s
V_9	Equilibrium time scour depth (t_e)	67–75 h
V_{10}	Experiment duration (t)	8 h

chamber on the downstream side and a sump tank for water flow at the upstream end for a steady flow rate.

The flume section has a working area that was 5 m long, 1.10 m wide, and 0.80 m deep, and it was located 5 m downstream from the inlet of the flume. The working area has been filled with a uniformly graded sand bed (180 mm thickness) was laid on a slope of 1/250. The characteristic properties of the sand were mean particle size, $d_{50} = 1.75$ mm, uniformity coefficient, $C_u = 2$ and curvature coefficient, $C_c = 1$. The

Table 5
Description of the variables used in the proposed investigation to formulate GEP.

Variables	Description	Values
$\mathbf{P_1}$	Functional set	+, -, *,/, Sqrt
$\mathbf{P_2}$	Chromosomes Number	30
$\mathbf{P_3}$	Head Size	10
$\mathbf{P_4}$	Number of Genes	3
$\mathbf{P_5}$	Size of Gene	32
$\mathbf{P_6}$	Connecting Function	+
$\mathbf{P_7}$	Fitness Function	RMSE
$\mathbf{P_8}$	Inversion Rate	0.00546
$\mathbf{P_9}$	Mutation Rate	0.00138
$\mathbf{P_{10}}$	One-point recombination Rate	0.00277
$\mathbf{P_{11}}$	Two-point recombination Rate	0.00277
$\mathbf{P_{12}}$	Stumbling Mutation	0.00141
$\mathbf{P_{13}}$	Gene recombination Rate	0.00277
$\mathbf{P_{14}}$	Root insertion sequence transposition Rate	0.00546
$\mathbf{P_{15}}$	Gene Transportation Rate	0.00277
$\mathbf{P_{16}}$	Uniform Gene Recombination	0.00755

Table 6
Values of gene constants used in the proposed equation Eq.

Gene Number	Constant	Value
G1	C_1	4.61
G1	C_5	−6.17
G1	C_6	−8.44
G1	C_9	−7.83
G2	C_1	−7.48
G2	C_5	−8.14
G2	C_6	1.73
G2	C_9	1.97
G3	C_0	−8.70
G3	C_3	1.93

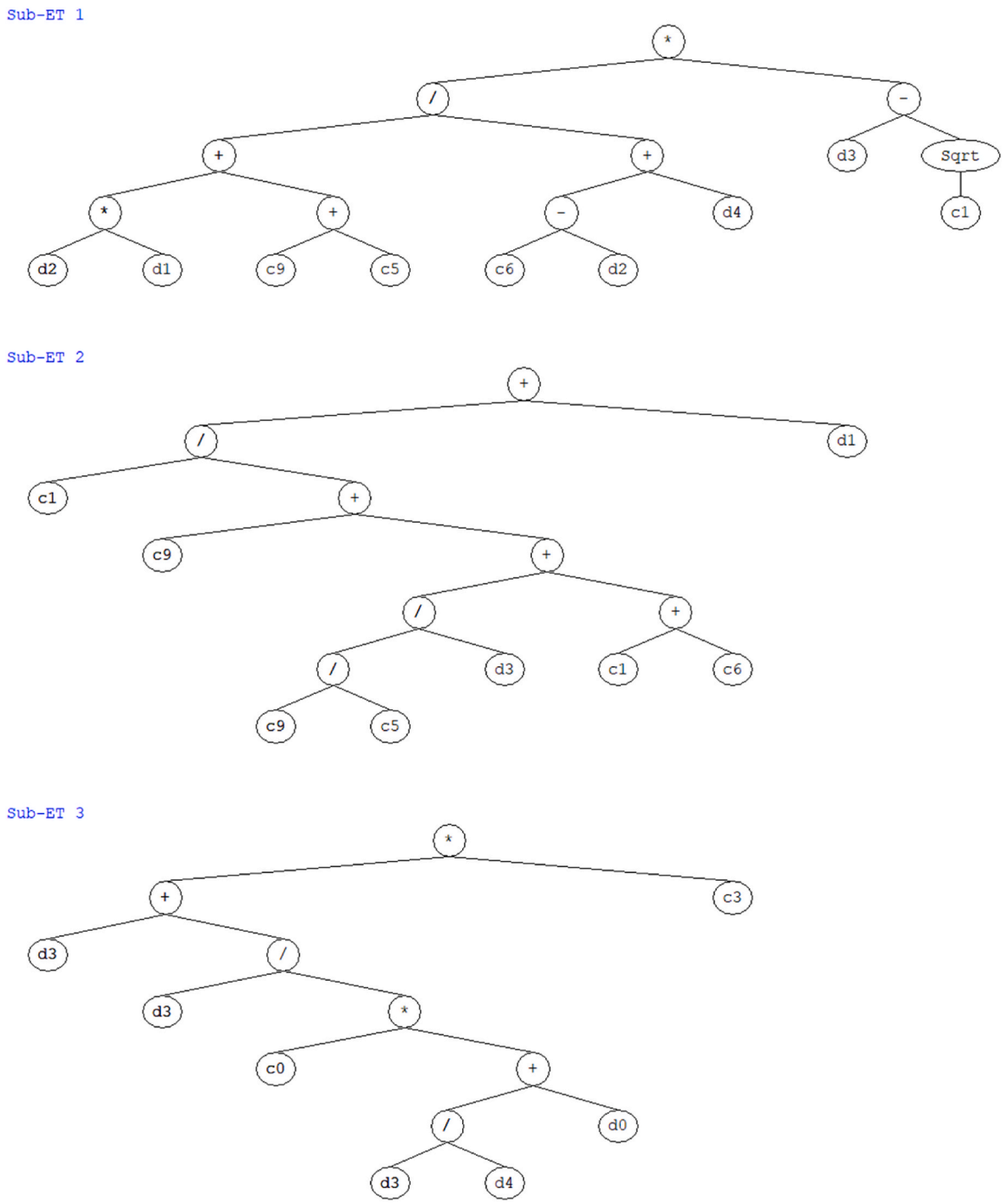


Fig. 5. Expression tree.

particle size distribution curve for the sediment used has been presented in Fig. 3. To study the flow structure around the bridge piers and reduce the flow shallowness effect depth of flow was adjusted between 0.12 and 0.163 m. As stated by Melville and Coleman (2000) that, if the ratio of the diameter of the pier to the flow depth (h/D) is less than 0.7 than flow shallowness effect can be neglected. The ratio of h/D for the present study is between 0.31 and 0.4, which is less than 0.7, so the flow shallowness effect is neglected in for the present study.

Acrylic pipes of diameter, *d*, 50 mm, has been used as pier models, laid in the center of the working section at pier spacing, $P_s = 1.5$, 2, 2.5, 3, 3.5, 4, 4.5, and *5* times the diameter of the pier. Pier models mounted with the ultrasonic transducers in the respective P_s (S1, S2, S3, S4, S5, S6, S7, and S8) are shown The whole experimental study has been carried out in two configurations (a) single pier, and (b) two piers in tandem positioning. Both the categories of experimental sets has been conducted for five rates of flow *(Q)* classified as A, B, C, D, and E, and eight P_s ratios. In this study, a wide range of pier spacing was adopted; the reason for this was that as the flow rate increases, a change in scour evolution pattern was observed. To analyze this process accurately, this wide range of pier spacing was adopted.

Forty-Five experiments have been performed using a single circular pier (5), and two piers (40), in tandem arrangement. The experimental run was conducted for 480 min. As concluded by Setia (2008), Yanmaz and Altinbilek (1991) and Mia et al. (2003), the MSD takes place in the

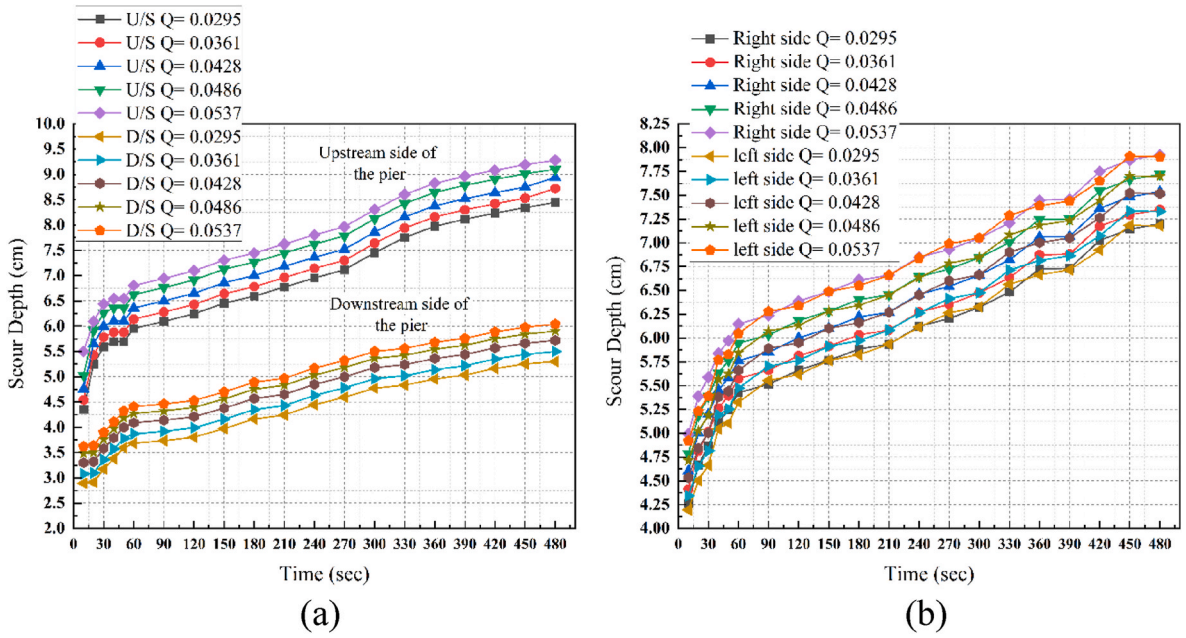


Fig. 6. Temporal scour evolution at the single pier.

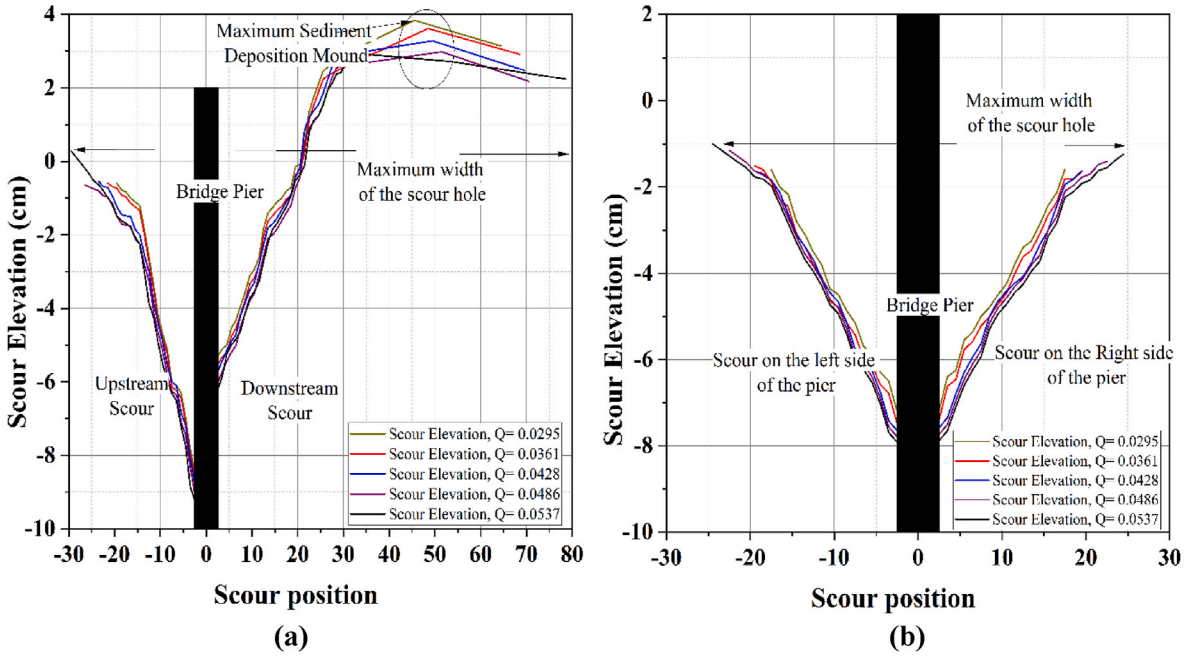


Fig. 7. Single pier scour hole on the (a) upstream and downstream side the pier; (b) left and right sides of the pier.

initial 180–240 min of the experimental run. Considering these references and keeping in view the existing laboratory facilities, the experiment duration for each run has been fixed as 480 min. All the experimental work has been carried out in clear water condition. For precisely measuring the pattern of the scour evolutions, bed measurements have been recorded with the SeaTek ultrasonic ranging transducers. Transducers data has been recorded after every 10 min for the first 60 min, and after every 30 min for the next 420 min of the experimental duration, for the eight observation points around the two piers. The water pump was turned off at the finish of each experimental test, and water was drained from the flume section. For final sand bed level measurements, a vernier-point gauge measurement with a precision of ±0.1 mm was utilized. The scour depth has been measured after 24 h of water drainage using a vernier point gauge with ±0.1 mm resolution, and an intense light was used to read the scour depth.

The SeaTek ultrasonic ranging system consists of 8 transducers, and works on the 5 MHz frequency. Each transducer has a diameter of 1 cm and was packed in stainless steel covering. The closest measurement range for these transducers was 3.5 cm, and the farthest measurement range was 110 cm. The transducers were numbered at both ends of the cable. When the cables were plugged into the appropriate connectors on the electronics package, the numbers correspond to the transducer numbers defined in the electronics. For the data recording from these transducers SeaTek software has been utilized. Eight transducers have been installed around both the piers as shown in Fig. 4.

Flow variables such as discharge, Froude number, flow head, the

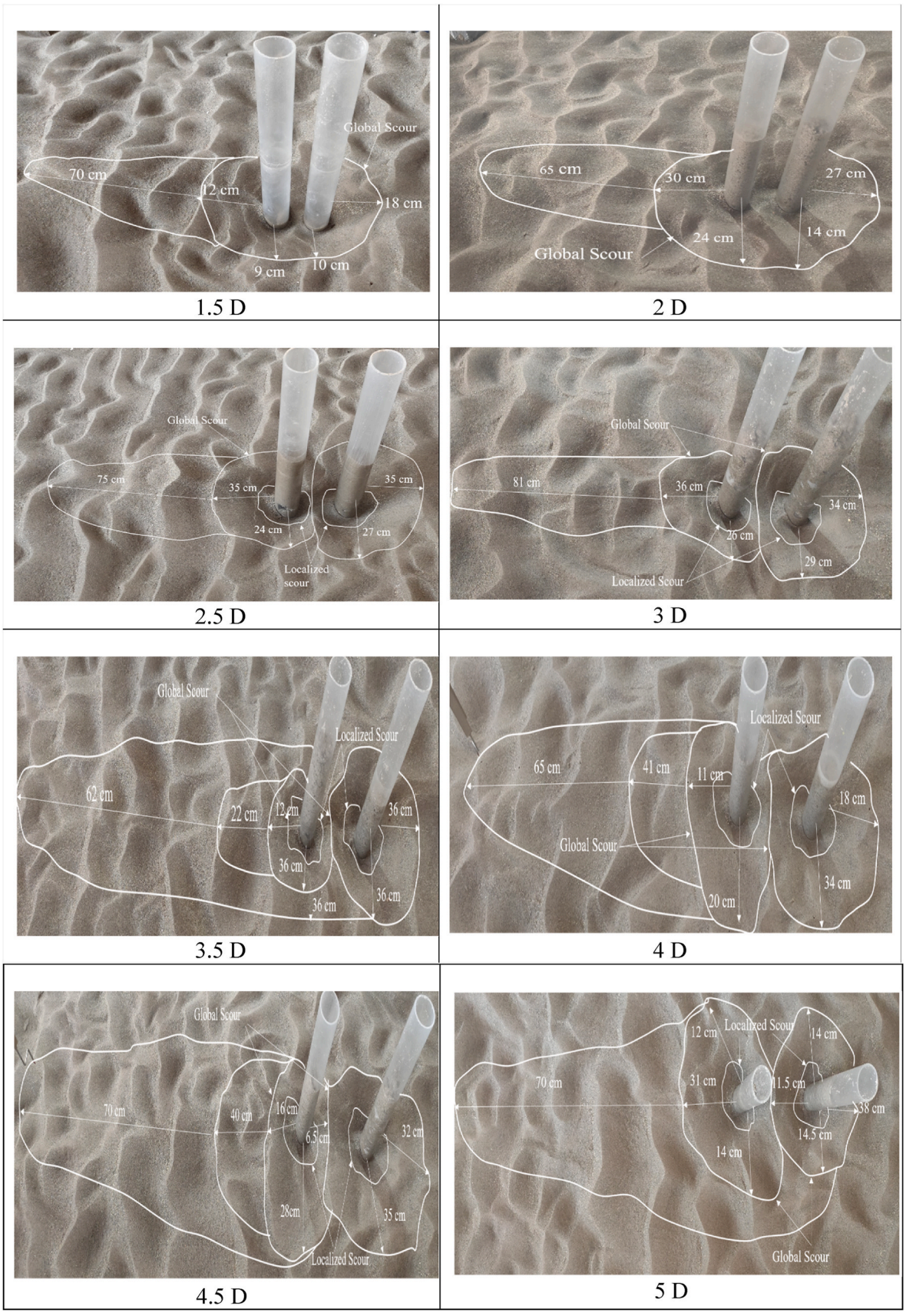


Fig. 8. Scour-hole dimensions for front and rear pier in tandem arrangement for maximum scour.

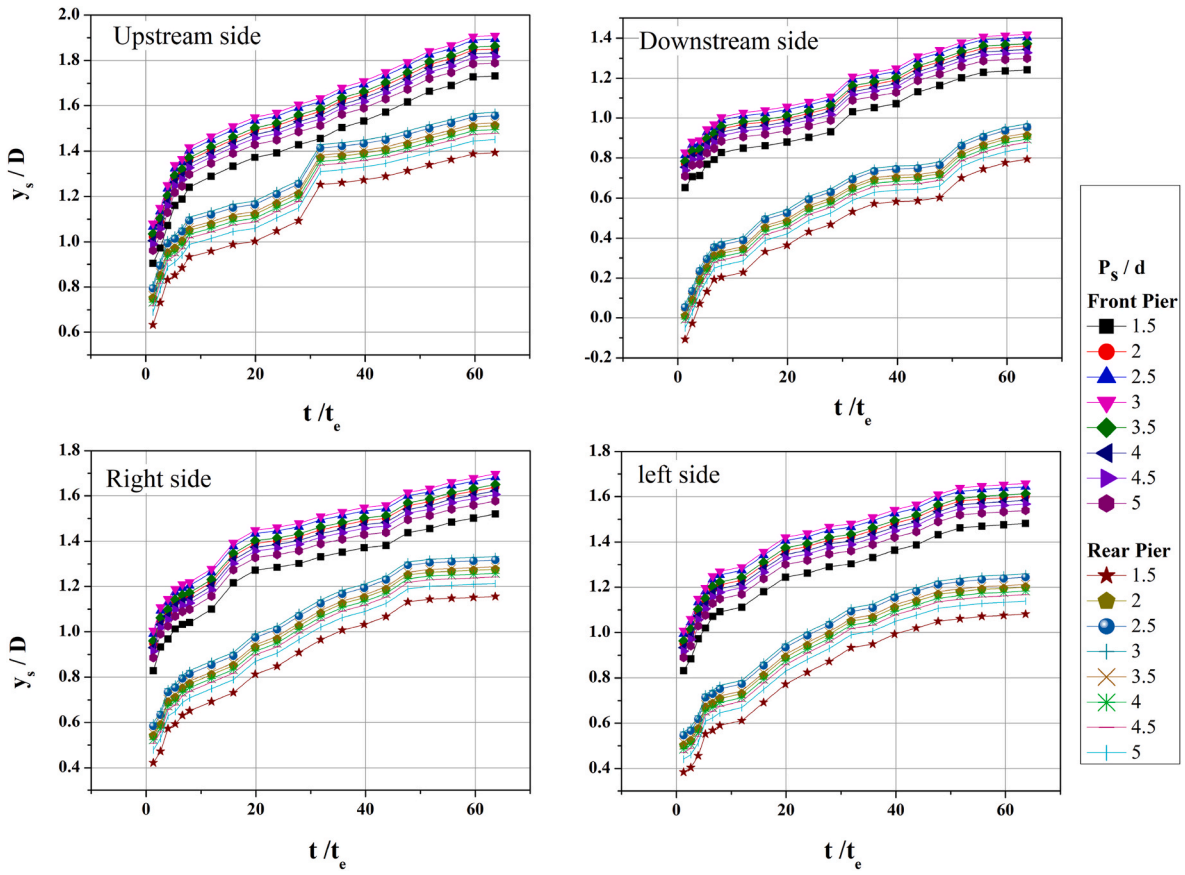


Fig. 9. Temporal scour evolution at two piers for rate of flow 0.0295 m^3/s.

velocity of flow, and critical velocity, for experimental conditions, has been presented in Table 2. Whereas the characteristic flow conditions for the present study and other similar studies are presented in Table 3.

4. Gene expression programming

A research technique that uses computer programmes is GEP (numerical articulations, decision trees, polynomial builds, and legitimate articulations). GEP combines both the advantageous aspects of GAs and GP approaches in order to overcome some of the limitations of AI. In 1999, Candida Ferreira developed the AI-based algorithm known as GEP (Ferreira, 2001). A linear chromosome, or expression tree, contains the genetic information of GEP models. ETs are fully advanced computational models that are typically designed to address a certain issue and are chosen based on how well they do so. The big discovery made by GEP was the development of chromosomes that could represent any expression tree. To read and express the information of GEP chromosomes, Ferreira (2001) developed a new language, which she called Karva language. Additionally, the architecture of chromosomes allowed for the production of many genes. No matter how much or how deeply we edit the chromosomes, the structural and functional arrangement of GEP genes always ensures the development of correct programmes. The genes are organized structurally into a head and a tail.

To begin, GEP generates a random population of separate chromosomes. Chromosomes have a fixed length and can contain one or more genes. The roulette wheel method selects the best chromosomes based on their fitness value. The selected chromosomes are then reproduced with some adjustments made by genetic operators; this procedure is repeated until the maximum number of generations is reached or the required accuracy is achieved. Mutation, Inversion (which reverses the sequence within the gene head), Insertions Sequence (IS) transport, Root Insertions Sequence (RIS) transport, Gene transport, Single or double crossover, and Gene crossover are the most common genetic operators used to modify chromosomes (Ferreira 2002; 2006).

4.1. GEP modelling

In the current study, the powerful software package GeneXproTools 5.0 was used to promote the development of GEP-based models for predicting the maximum scour depth that may occur at a bridge pier. Data sets from Keshavarzi et al. (2018); Liang et al. (2017); Malik and Setia (2020, 2018) were collected to develop the GEP model for predicting local scouring around a bridge pier.. These data were then fed to the GEP in the form of a dependent variable (Y_s/d) and independent input variables (F_r, U/U_c, d/d_{50}, t/t_e, P_s/d), with the output variable Y_s/d then developed using the GEP.

The GEP randomly divided the data into validation and testing data (20%) and training data (80%), with the training data being used to build the GEP model. The GEP parameters and operations were then defined in 6 steps to stimulate the generation of the mathematical equation needed to predict local scouring around a bridge piers.

1. To begin with GEP modeling first step is generating an initial population. As suggested by (Ferreira, 2001) any number of population can be utilized in the underlying populace but population of 30–100 offers optimal results.

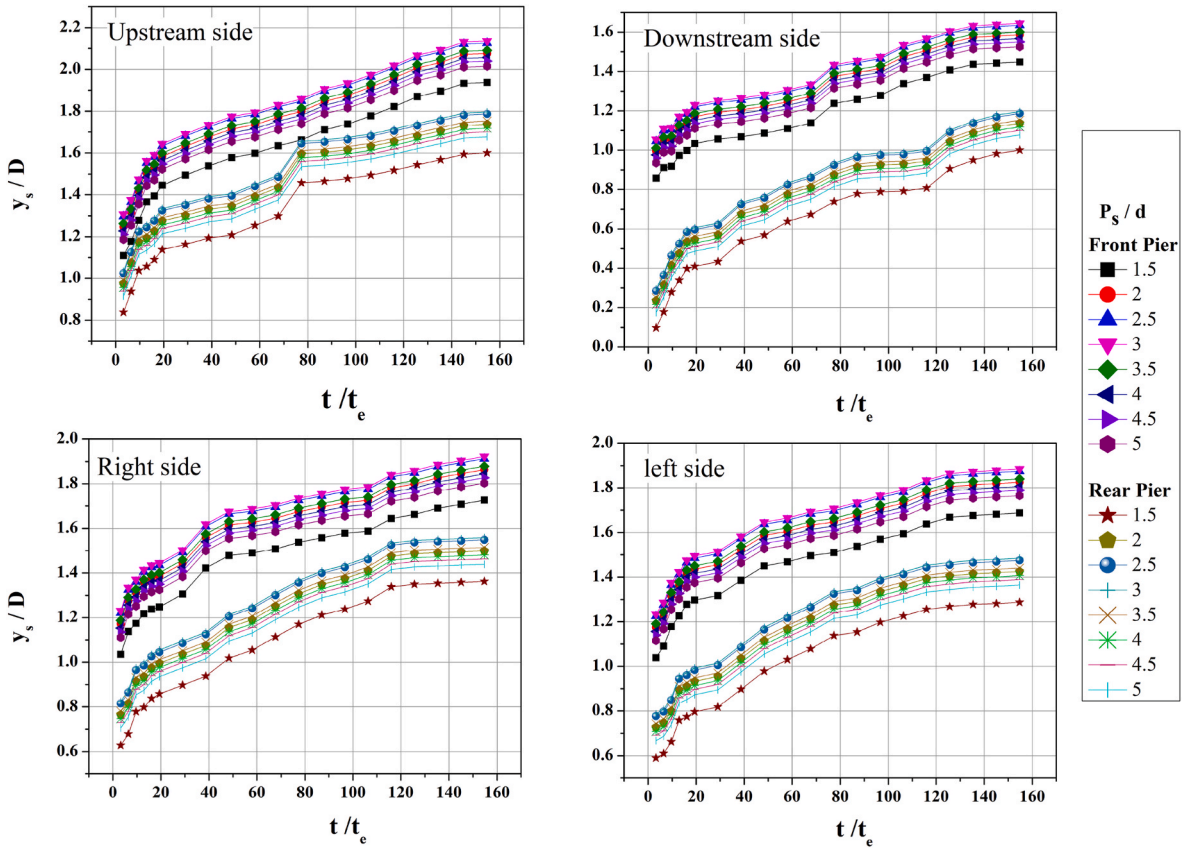


Fig. 10. Temporal scour evolution at two piers for rate of flow 0.0361 m^3/s.

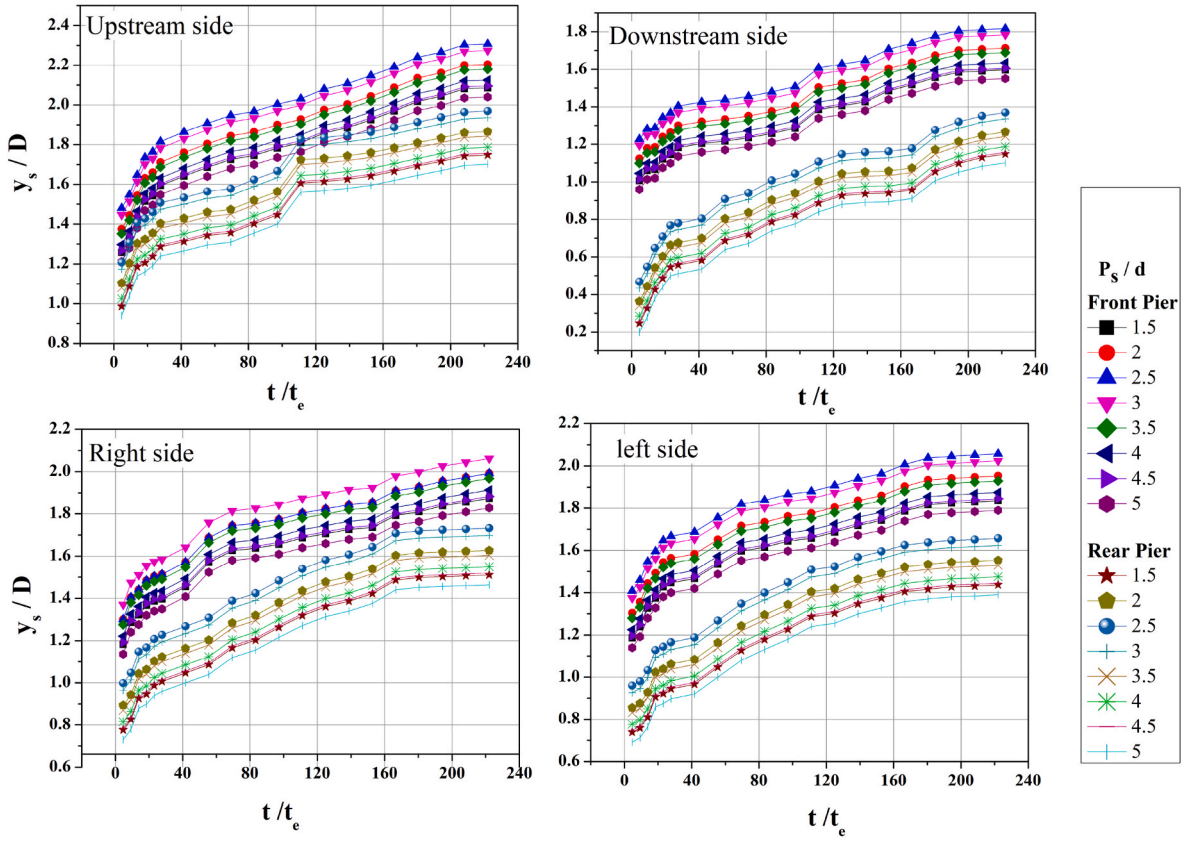


Fig. 11. Temporal scour evolution at two piers for rate of flow 0.0428 m^3/s.

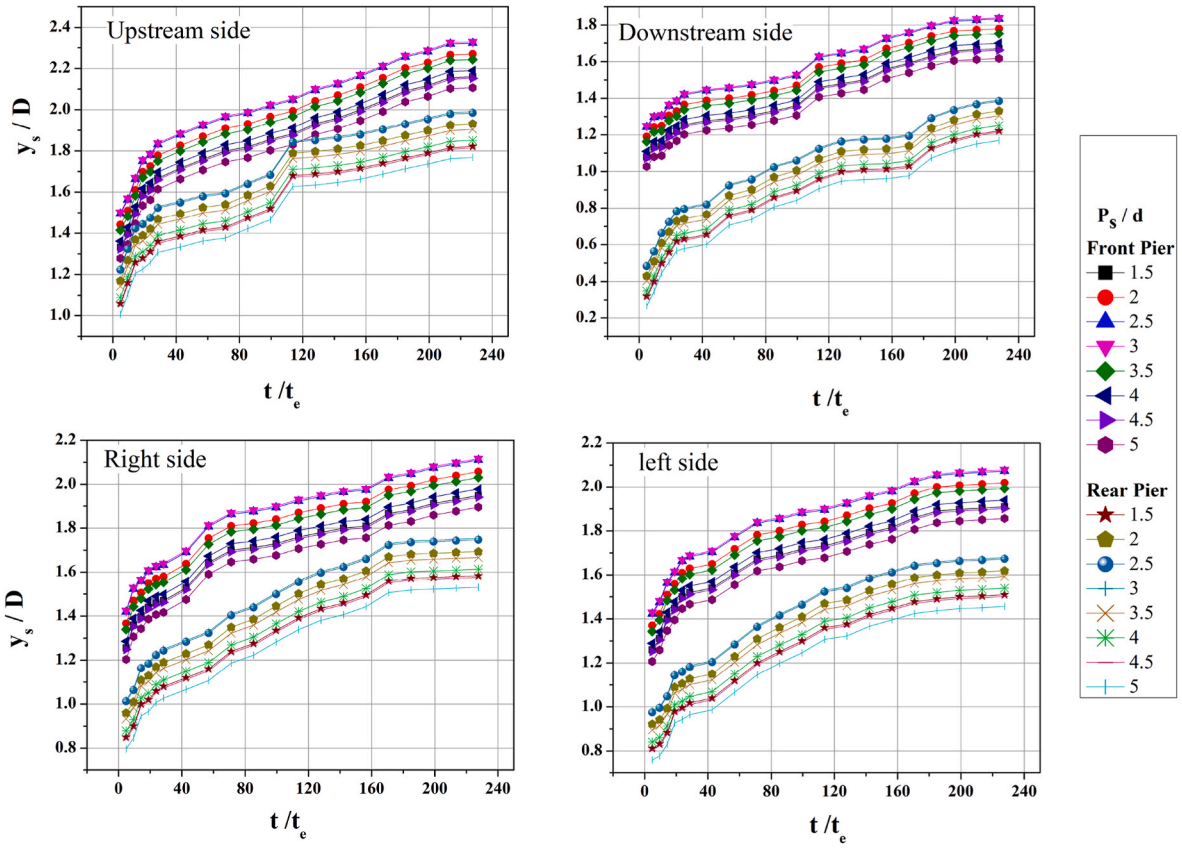


Fig. 12. Temporal scour evolution at two piers for rate of flow 0.0486 m^3/s.

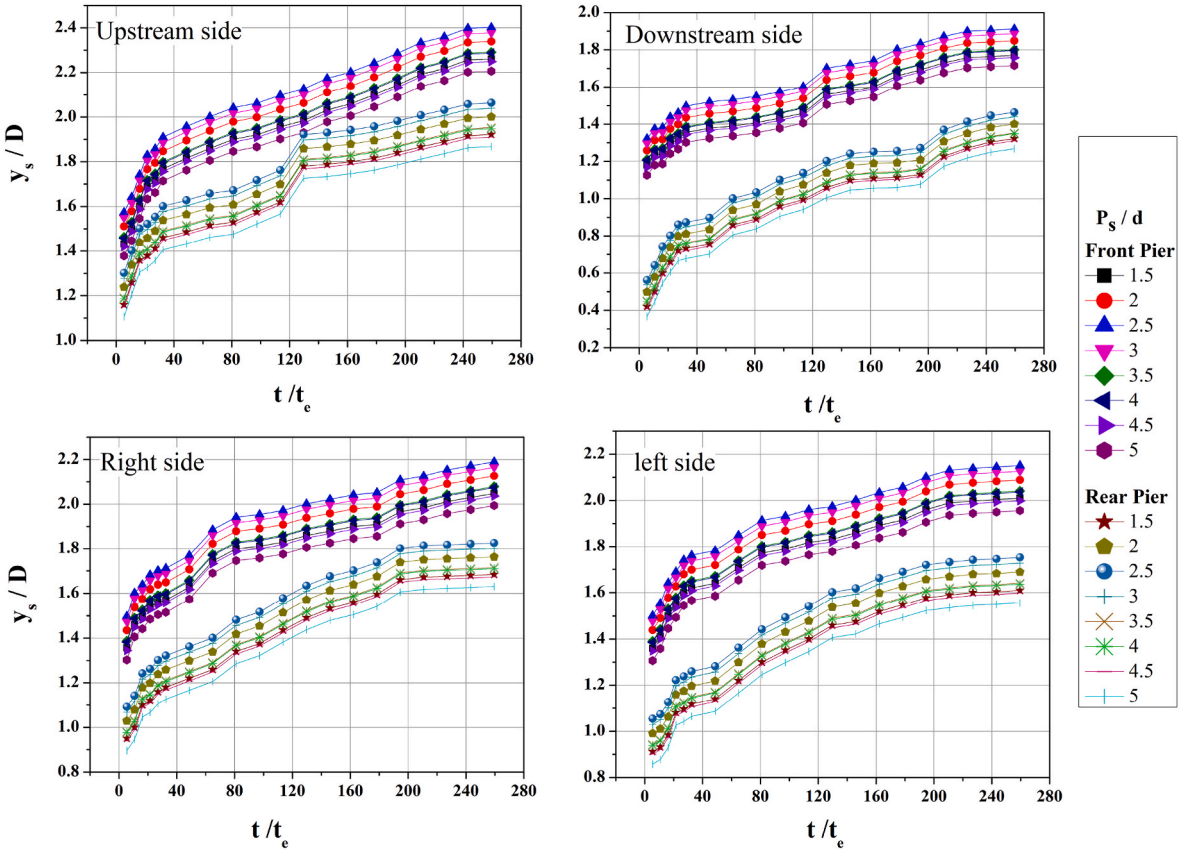


Fig. 13. Temporal scour evolution at two piers for rate of flow 0.0537 m^3/s.

2. After instating the population size, the fitness function is measured for measurement of fitness function for each individual chromosome by assessing RMSE.
3. After choosing the RMSE as fitness factor, the following stage is to select the function and terminals for each chromosomes. Functional series and terminals used in the present study are (+, -, *,/, sqrt) and (F_r, U/U_c, d/d_{50}, t/t_e, P_s/d).
4. Next advance is managing the architecture of the chromosomes by choosing the Gene numbers and length of head and tail for every gene.
5. The connecting function between the sub expression trees (sub-ETs) was chosen in the following stage.
6. Following the preceding stage, genetic operators for chromosome design are chosen. These genetic operators include transposition (gene transposition, IS, and RIS), mutation (gene recombination, both one-point and two-point), inversion, and Dc-specific genetic operations and its rate.

After determining all genetic parameters, the model was simulated for a number of generations in excess of 50,00,000. The values of all the parameters used in the present study are tabulated in Table 4, Table 5 and Table 6. The proposed scour depth (Y_s/d) equation is represented in the form of expression tree (ET) in Fig. 5. The ET's expression can be written in the simplified form as:

5. Results and discussion

The following results have been presented in the current study: (1) temporal evolution of scour depth around the piers, (2) effect of densimetric Froude number on MSD, (3) effect of u/u_c on MSD, (4) comparison of the existing equations with the proposed GEP equation. Each section is separately examined for both the experimental categories (single pier and two piers). The results of each subsection are presented at the end of each section.

6. Temporal evolution of scour depth

Fig. 6 presents that, the temporal evolution of scour depth for an increasing rate of flow around the single pier. The value of Y_s increases rapidly during the first hour of experimental duration, and after that, it tends to grow at a delayed rate until it reaches an equilibrium state. After 360 min of the experimental run, the rate of temporal evolution becomes approximately constant on the upstream and downstream side of the pier. However, on the right and left side of the pier, the equilibrium stage was achieved after a lapse of 420 min, additionally, a constant rate was also observed at 360 min, but within the next 30 min, scour depth again started increasing and attained equilibrium at 420 min.

In the equilibrium state, the scour hole around the single pier is depicted in Fig. 7. The maximum scour depth was observed at upstream of the pier. The scoured region extends into the region surrounding the

$$\frac{Y_s}{D}=\left\{\frac{(d_2\times d_1)+(C_9+C_5)}{(C_6-d_2)+d_4}\times\left(d_3-\sqrt{C_1}\right)\right\}+\left\{\frac{C_1}{\left[\frac{C_9/C_5}{d_3}+(C_1+C_6)\right]+C_9}+d_1\right\}+\left\{\left[\frac{d_3}{\left(\frac{d_3}{d_4}+d_0\right)\times C_0}+d_3\right]\times C_3\right\} \tag{5}$$

By substituting the values of variables equation can be written as:

$$\frac{Y_s}{D}=\left\{\frac{\left(P_s/D\times U/U_c\right)+(C_9+C_5)}{(C_6-P_s/D)+D/d_{50}}\times\left(t/t_e-\sqrt{C_1}\right)\right\}+\left\{\frac{C_1}{\left[\frac{C_9/C_5}{t/t_e}+(C_1+C_6)\right]+C_9}+U/U_c\right\}+\left\{\left[\frac{t/t_e}{\left(\frac{t/t_e}{D/d_{50}}+F_r\right)\times C_0}+t/t_e\right]\times C_3\right\} \tag{6}$$

Further by substituting the values of the coefficients and simplifying the proposed equation can be written as:

$$\frac{Y_s}{D}=\left\{\frac{\left(P_s/D\times U/U_c\right)-14}{\left(D/d_{50}-P_s/D-8.45\right)}\times\left(t/t_e-2.14\right)\right\}+\left\{\frac{7.48}{\left[\frac{0.242}{t/t_e}+3.78\right]}+U/U_c\right\}+\left\{\left[\frac{t/t_e}{\left(\frac{t/t_e}{D/d_{50}}+F_r\right)\times(-8.70)}+t/t_e\right]\times 1.93\right\} \tag{7}$$

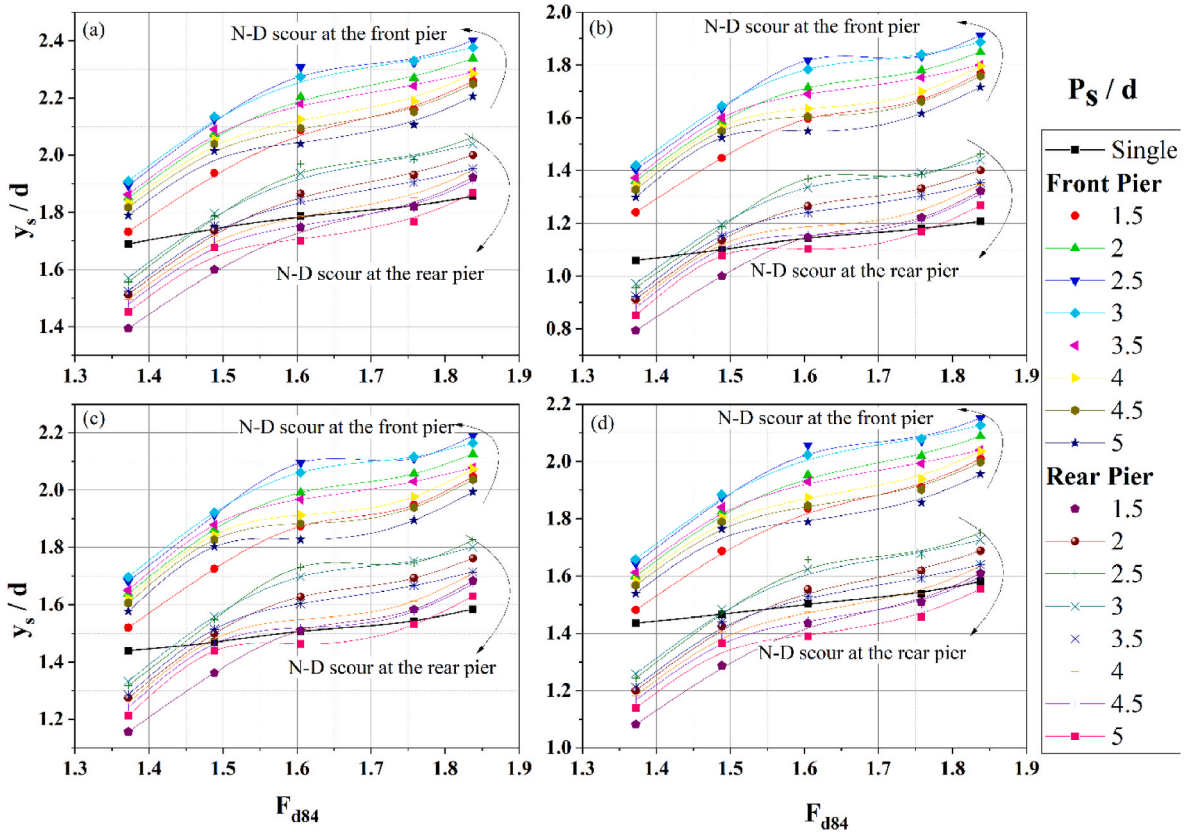


Fig. 14. Effect of densimetric Froude number on non-dimensional scour depth at front and rear pier at (a) upstream side (b) downstream side (c) right side and (d) left side.

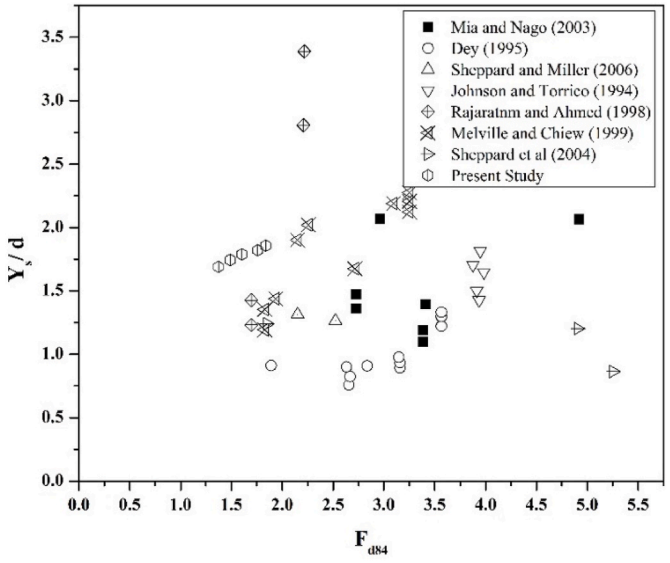


Fig. 15. Non-dimensional scour depth variation with Densimetric Froude number (F_d).

pier, and the scoured sediment was deposited on the downstream side (behind) of the pier. Fig. 4a and b present erosion and deposition of sediments around the pier. The scour at the left and right observation points of the cylinder show similar trend. It can be concluded from Fig. 6, that 81.5% of equilibrium scour depth developed within 360 min of experimental duration. It also demonstrates the bed surface elevation and clearly indicates the scour hole development with the time around the single pier. Moreover, after 75% of the equilibrium scour time, the rate of increase in scour depth will almost flatten. The scour depth immediately in front of the cylinder was considered as maximum scour depth which was 87.4, 85.8, 84.2, 81.6, and 80.4 percentage of the equilibrium scour depth after 480 min, calculated by using the equations Melville and Chiew (1999), written in Eq. (8) and (9). Despite the fact that various equations have been developed to estimate the time of equilibrium scour depth development,(Melville and Chiew, 1999), equations have been observed to be very practical and simple, with the ability to predict a reasonably precise outcome. As a result, the current study employs these equations for additional analysis and calculation of temporal development scour depth around bridge piers.

$$Y_{t_e} = \exp\left\{ -0.03 \left| \frac{u_c}{u} \ln\left(\frac{t}{t_e}\right) \right|^{1.6} \right\} \tag{8}$$

where t_e is given as:

$$t_e\ (\text{days}) = 30.89 \frac{d}{u}\left(\frac{u}{u_c} - 0.4\right)\left(\frac{h}{d}\right)^{0.25} \quad \text{for } \frac{h}{d} < 6, \frac{u}{u_c} > 0.4 \tag{9}$$

where, Y_{t_e} = scour depth percentage attained at equilibrium time; u_c = approaching flow critical velocity; t = duration of the experiment; t_e = time needed to achieve equilibrium scour depth; d = diameter of the pier; h = depth of flow.

The flow field surrounding the piers altered significantly as water flowed through the working area, causing a downward flow. The downstream flow eroded the sediments at the upstream surface, and the separated shear layer rolled upward, forming horseshoe vortices. The fast flow on the piers side caused the range of scour to expand due to the influence of streamline contraction, which led to the rapid development of scour in the first 60 min. The horseshoe vortices and vortex shedding flow around the piers continued to erode the sediments after 60 min,

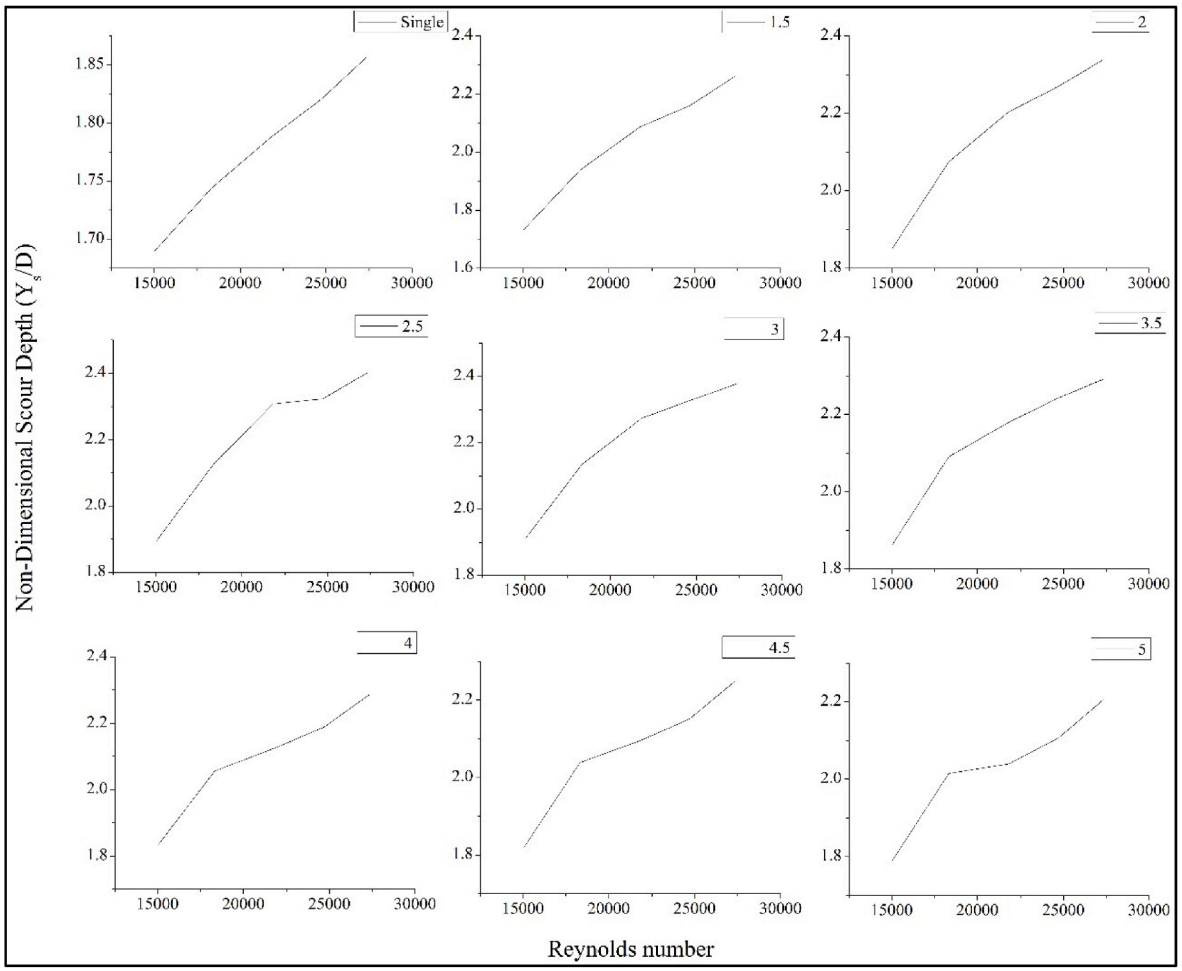


Fig. 16. Reynolds number effect on the non-dimensional scour depth for the front pier.

though at a slower rate than at the commencement of the scour process. Sediments were constantly moved downstream, causing the scour depth to gradually grow over the next 300 min. However, as compared to the first stage of scour, the scour rate was reduced. After 300 min, the shear tensions at the bed surface induced by flow and vortices were no longer adequate to promote continual erosion of the seabed sediments, and scour attained its 90% equilibrium state.

Fig. 8 depicts the scour hole dimensions for both the pier in the tandem arrangement, scour holes around both the piers. As depicted in Fig. 8 the local and global scour is observed around the piers in case of P_s 1.5D- 5D. For small P_s = (1.5–2)D, only global scour hole was observed, for the P_s =(2.5–5)D both localized and global both the scour holes were observed. This is because of the strong scouring intervention at individual piers, which removes out the sediments between the piers. As the P_s increases localized scour hole at individual piers emerge as the waning interference of scouring between individual piers. Scour hole interference increases while the localized scour hole dominate the global scour at the piers as a result hydraulic group effect was vanished. The down flow erodes the sediments surrounding the front pier, wake vortices began to form around the rear pier for the small pier spacing (1.5D-2D) front pier and the scour hole development around the pier is in the shape of a single scour hole. However, due to the large pier spacing (2.5D-5D), the development of two independent scour holes commenced, as did the formation of horseshoe vortices around the rear pier.

Non-dimensional temporal scour depth for two piers under five rate of flow conditions are presented in Fig. 9, Fig. 10, Fig. 11, Fig. 12, and Fig. 13, respectively. A detailed comparison of the scour depth evolution at all the observation points (upstream, downstream, right and left side of the pier) is presented here. As shown in the upstream curves, the scour depth at the upstream face of the front pier was greater than the rear pier. Similarly, the scour rate on the left and right side of the front pier was greater, than the rear pier. Therefore, this can be also said that the scour pattern at both piers was not identical. The MSD was monitored at the upstream side of the front pier and minimum at the downstream side of the rear pier. Hence the maximum scour depth was observed at P_s/d = 2.5, which was 24% greater than the scour at a single pier, and the minimum scour depth was observed at the P_s/d ratio 5, which was 17% less than the scour depth at the front pier. As non-dimensional pier spacing (P_s/d) was increased, the extent of scour depth decreased until P_s/d ratio 3, after that scour started evolving, and piers started behaving individually. Up to P_s/d ratio 2.5, the scour hole developed was acting like a single body; however, after P_s/d ratio 3–5, the formation of two separate scour holes was witnessed. Scour around the front pier was similar as around the isolated pier, similar scour behaviour was observed by Wang et al. (2016a) in his research work. Scour around the front pier was nearly the same for multiple varying P_s as it was around an isolated pier. So, we can conclude that the maximum influence of the front pier on the rear pier was predominating up to P_s/d ratio equal to 3; after that it starts reducing.

7. Densimetric Froude number

The densimetric Froude number is essential in determining the MSD around bridge piers.in the present study, densimetric Froude number (F_d) has been used instead of Froude number (F_r). According toCheng

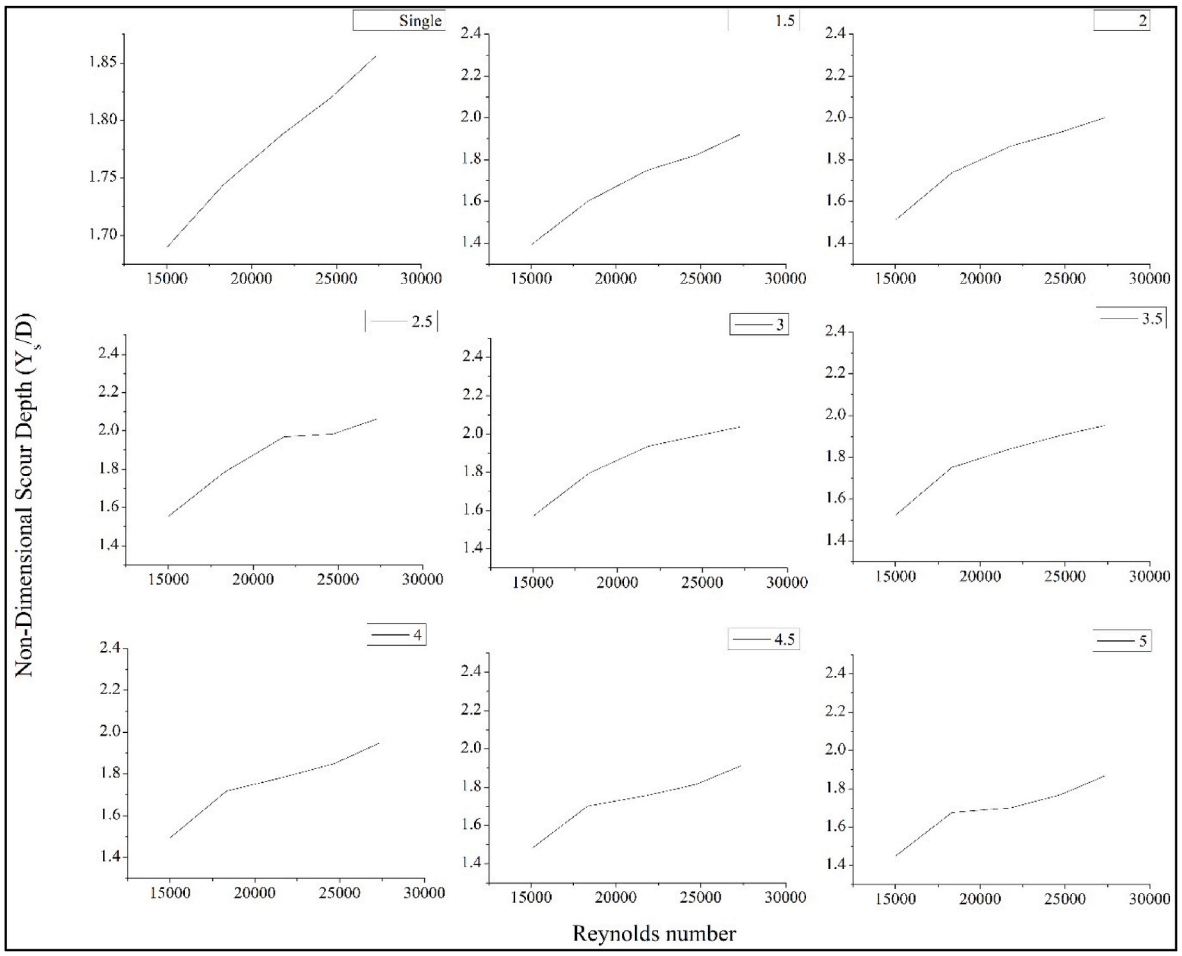


Fig. 17. Reynolds number effect on the non-dimensional scour depth for the rear pier.

et al. (2020) the size of representative bed particles cannot be described using only the median diameter d_{50} in non-uniformly distributed sediments, and conducted a study using the different particle diameters, including d_{16}, d_{35}, d_{50}, d_{65}, d_{84}, d_{90}, d_{95}, and d_{99}.study results shows that the, the best results are obtained by using the d_{84} particle size, so in the present study d_{84} particle has been utilized rather than d_{50}. The expression proposed by has been used to compute the densimetric Froude number is $F_d = \frac{u^2}{gd_{84}}$, where d_{84} is the representative size of the sediment material is given by Cheng et al. (2020). The values of the sediment size d_{84} is interpolated from the sieve analysis curve, presented in Fig. 3, and can be written as in Eq. (10):

$$y = y_o + \frac{A}{w\sqrt{\pi/2}} e^{-2\frac{(x-x_c)^2}{w^2}} \tag{10}$$

The results of the present study are illustrated in Fig. 14, the effect of densimetric Froude number on non-dimensional scours depth with non-dimensional pier spacing P_s/d. Fig. 14 shows that, F_d ranges from 1.4 to 1.8, and as the value of the Froude number increases, the MSD also increases. For all the flow conditions, the same range of Froude number was adopted to evaluate the change in the MSD evolution pattern. The maximum non-dimensional scour depth was 2.4, observed at the upstream of the front pier for $P_s/d = 2.5$, at the highest value of the densimetric Froude number $F_{d84} = 1.83$. And the minimum non-dimensional scour depth was 1.5, observed at the downstream side of the rear pier at the lowest value of the $F_{d84} = 1.37$, which was 21.6% less than that of MSD at the upstream side of a single pier. As the value of densimetric Froude number increases, non-dimensional scour depth also increases. For the value of $F_{d84} \leq 1.6$, the minimum scour depth was observed for the $P_s/d = 1.5$, and for the $F_{d84} > 1.6$, the minimum scour depth was observed in the case of $P_s/d = 5$. The MSD for $F_{d84} = 1.4$ -1.8 occurs at the $P_s/d = 2.5$ at the upstream side of the front pier. As a result, the bridge scour was more sensitive on the upstream side of the front pier.

Furthermore, the findings of the current study were compared to previously publish experimental results. The experimental results of the Mia and Nago (2003), Dey et al. (1995), Sheppard and Miller Jr. William (2006), Johnson and Torrico (1994), Ahmed and Rajaratnam (1998), Melville and Chiew (1999), Jones and Sheppard (2004) and the present study are presented in Fig. 15. This section shows the non-dimensional scour depth variation with the densimetric Froude number. As a result of Fig. 15, it is possible to conclude that the findings of the current study are in good agreement with the available literature. The experimental investigation concludes that, the pier spacing and densimetric Froude number play an important role in evaluating the MSD, and that these factors must be incorporated in regression method for accurate prediction of the MSD.

8. Effect of Reynolds number on maximum scour depth

Effect of Reynolds number on the non-dimensional scour depth has been studied with the variation on the pier spacing. The Reynolds number based on the depth-averaged approach velocity (U) and the diameter of the obstacle (d), $R_e = Ud/\upsilon$, where υ is the kinematic viscosity of the fluid. Fig. 16 and 17 depicts the variation in the non-dimensional scour depth for the front and rear pier. As shown in the Fig. 16, for the single pier the Y_s/d curve is linearly varying with the increase in the Reynolds number. As the P_s/d is between 1.5 and 3 then

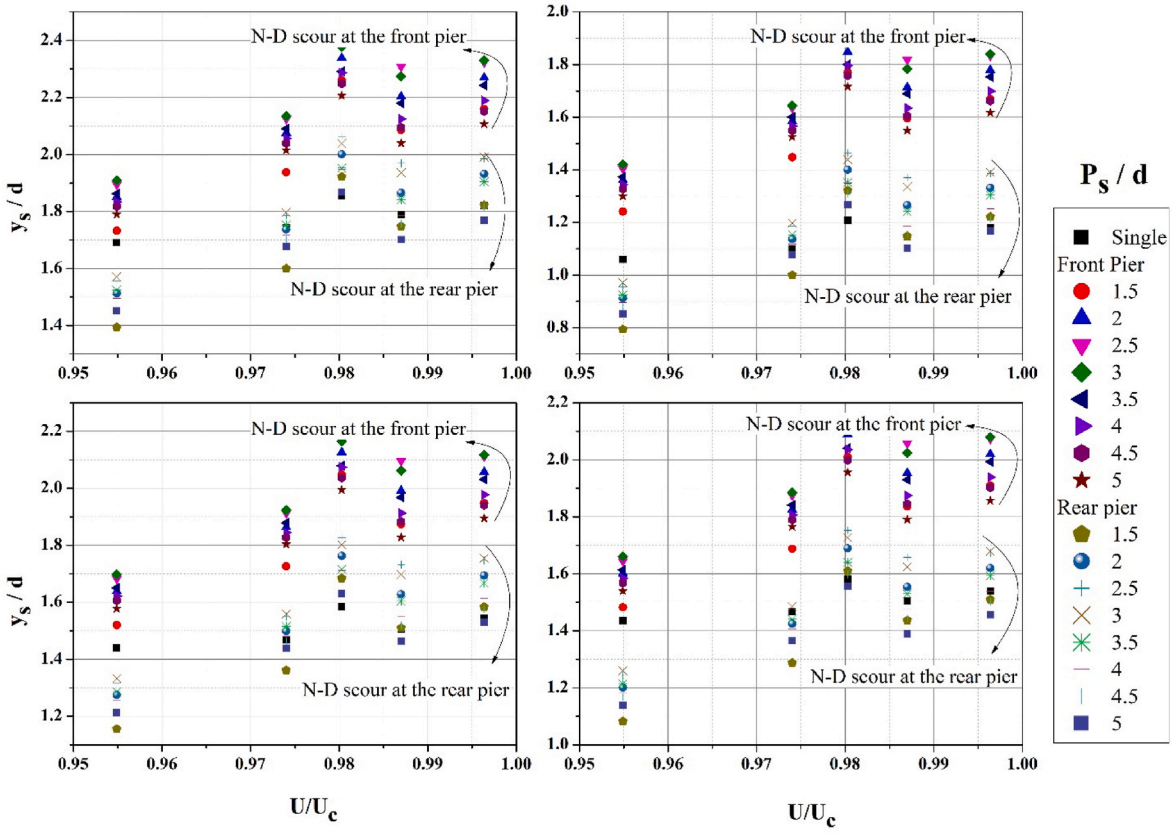


Fig. 18. Non-dimensional scour depth (y_s/d), versus relative velocity (u/u_c), with the increasing non-dimensional pier spacing (P_s/d) at (a) upstream side (b) downstream side (c) right side and (d) left side.

Table 7
Experimental results of first category (Single pier) for all the five flow conditions.

Run No.	Y_{SU} (cm)	Y_{SR} (cm)	Y_{SL} (cm)	Y_{SD} (cm)
A	8.45	7.2	7.18	5.3
B	8.72	7.34	7.33	5.5
C	8.94	7.53	7.52	5.72
D	9.10	7.72	7.69	5.90
E	9.28	7.92	7.90	6.04

where, Y_{SU} = scour depth at the upstream side of the pier; Y_{SR} = scour depth at the right side of the pier; Y_{SL} = scour depth at the left side of the pier; Y_{SD} = scour depth at the downstream side of the pier.

pier has been increased the Y_s/d curve slightly flattened on the higher side of the curve, whereas for the P_s/d 3–4.5 the curve is flattened more on the higher side. The maximum change in the individual curves have been observed for the R_e = 17,000–27,000. For the P_s/d is between 1.5 and 3, at R_e = 17,000, slight increase in the non-dimension scour have been obtained, as the value of the R_e = 22,000. For the P_s/d 3–4 the Y_s/d curve flatted for the range of R_e 17,000–25,000 and for the P_s/d = 4.5–5 the Y_s/d curve attains the least values for the R_e 20,000–27,000. This fattening of the curve is the resultant increased pressure gradients induced by the piers trigger the separation of the incoming flow and involvement of the wake vortices with the horseshoe vortices in the pier spacing area.

9. Effect of relative velocity on maximum scour depth

In this section, the effect of relative velocity (u/u_c) on the non-dimensional scour depth (Y_s/d) has been presented for the non-dimensional pier spacing (P_s/d); results are presented in Fig. 18. In the first category of the experimental study, the rationality of the study is demonstrated for the single pier in Table 7. The experimental results of the current study agree closely with the Eq. proposed by (Aguirre-Pe et al., 2003) and also used by authors in studies (Wang et al., 2016a, 2016b), is written as in Eq. (8)

$$\frac{Y_s}{d} = 2.865\frac{u}{u_c} - 1.332 \tag{11}$$

Where, Y_s = scour depth; d = pier diameter; u = averaged velocity of approaching flow; and u_c = critical velocity of the approaching flow.

The coefficient in Eq. (8) is slightly less than in Chiew's equation. Because the duration of the experiments used in the present study is somewhat less than that taken to obtain real-time equilibrium, and the physical properties of the sediment used in the study vary considerably with those of river sand used by chiew in his research (Chiew, 1995; Qi et al., 2013). The mean value of critical velocity observed in the present study agrees well with the Schamovis equation (Tang et al., 2013), earlier used by authors in their research work (Wang et al., 2016a, 2016b). Schamovi's equation can be written as in Eq. (12):

$$u_c = 1.14\sqrt{gd_{50}\frac{\rho_s - \rho}{\rho}}\left(\frac{h}{d_{50}}\right)^{1/6} \tag{12}$$

where, ρ = specific gravity of the water; ρ_s = specific gravity of the sediment; g = acceleration due to gravity; and d_{50} = mean particle size.

In the second category of the experimental programmes, two pier in the tandem position were considered with the increasing P_s between the piers. The results of the experimental study, are shown in Table 8. The results show that, at P_s equals to zero, both the piers behave like a single body of different shapes and diameters, this case can be treated the same as a single pier. Piers act like a one body because of smaller P_s between

Table 8
Experimental results of second category (two pier) for all the five flow conditions.

	Front Pier				Rear Pier			
Run No	Y_{SU} (cm)	Y_{SR} (cm)	Y_{SL} (cm)	Y_{SD} (cm)	Y_{SU} (cm)	Y_{SR} (cm)	Y_{SL} (cm)	Y_{SD} (cm)
AS1	8.66	7.6	7.41	6.21	6.97	5.78	5.41	3.97
AS2	9.25	8.19	8.00	6.80	7.56	6.37	6.00	4.56
AS3	9.47	8.41	8.22	7.02	7.78	6.59	6.22	4.78
AS4	9.55	8.48	8.29	7.09	7.85	6.66	6.29	4.85
AS5	9.31	8.25	8.06	6.86	7.62	6.43	6.06	4.62
AS6	9.17	8.11	7.92	6.72	7.48	6.28	5.92	4.48
AS7	9.08	8.03	7.84	6.64	7.39	6.21	5.84	4.39
AS8	8.95	7.88	7.69	6.49	7.25	6.06	5.69	4.25
BS1	9.69	8.63	8.44	7.24	7.99	6.81	6.44	4.99
BS2	10.37	9.32	9.13	7.92	8.68	7.49	7.13	5.68
BS3	10.62	9.56	9.37	8.17	8.93	7.74	7.37	5.93
BS4	10.67	9.61	9.42	8.22	8.98	7.79	7.42	5.98
BS5	10.45	9.39	9.20	8.00	8.76	7.57	7.20	5.76
BS6	10.28	9.22	9.03	7.83	8.59	7.40	7.03	5.59
BS7	10.19	9.14	8.95	7.75	8.51	7.32	6.94	5.51
BS8	10.07	9.01	8.82	7.62	8.38	7.19	6.82	5.38
CS1	10.43	9.37	9.18	7.98	8.74	7.55	7.18	5.74
CS2	11.02	9.96	9.77	8.57	9.33	8.14	7.77	6.33
CS3	11.54	10.47	10.29	9.09	9.85	8.66	8.28	6.85
CS4	11.37	10.31	10.12	8.92	9.68	8.49	8.12	6.68
CS5	10.89	9.84	9.65	8.47	9.21	8.02	7.65	6.21
CS6	10.62	9.56	9.37	8.17	8.94	7.74	7.37	5.93
CS7	10.47	9.42	9.23	8.03	8.78	7.59	7.23	5.78
CS8	10.19	9.14	8.95	7.75	8.51	7.32	6.95	5.51
DS1	10.79	9.74	9.55	8.35	9.12	7.92	7.55	6.11
DS2	11.35	10.28	10.09	8.89	9.66	8.47	8.09	6.66
DS3	11.62	10.55	10.36	9.16	9.93	8.73	8.36	6.92
DS4	11.65	10.58	10.39	9.19	9.96	8.76	8.39	6.95
DS5	11.22	10.15	9.96	8.77	9.52	8.34	7.96	6.52
DS6	10.95	9.88	9.69	8.49	9.26	8.06	7.69	6.25
DS7	10.76	9.70	9.51	8.31	9.07	7.88	7.51	6.07
DS8	10.53	9.47	9.28	8.08	8.84	7.65	7.28	5.84
ES1	11.30	10.24	10.05	8.85	9.61	8.42	8.05	6.61
ES2	11.69	10.63	10.44	9.24	10.0	8.81	8.44	7.00
ES3	12.01	10.95	10.76	9.56	10.32	9.13	8.76	7.32
ES4	11.88	10.83	10.64	9.44	10.19	9.01	8.64	7.19
ES5	11.45	10.39	10.20	9.00	9.76	8.57	8.20	6.76
ES6	11.43	10.37	10.18	8.98	9.74	8.55	8.18	6.74
ES7	11.24	10.18	9.992	8.79	9.55	8.36	7.99	6.55
ES8	11.03	9.97	9.78	8.58	9.34	8.15	7.78	6.34

the piers, the front pier exerts the sheltering effect on the rear pier. Because of this sheltering effect, the upstream nose of the front pier behaves like a single pier, and there was no scour effect on the rear pier; similar results were obtained in their research by Breusers et al. (1977) and Shen et al. (1969).

As the value of P_s/d between the piers increased, the relationship between the y_s/d and u/u_c for five flow conditions agrees well with the Eq. (8) as shown in Fig. 18. The findings of the present study illustrate that the scour hole developed at the front pier was larger in size than that of the rear pier and single pier; these results were similar with the results obtained by Hannah (1978) and Behzad and Beheshti (2006).

9.1. GEP model performance analysis for equilibrium scour depth

A GEP formulation has proposed in the present study is written in Eq. (7). The performance analysis of the proposed equation has been carried out by comparing the GEP equation with the well-known existing equations, IRC equation (Congress, 2016), Larsen and Toch equation (Laursen and Arthur, 1956), Melville and Coleman equation (Melville and Coleman, 2000) and HEC-18 equation (Richardson and Davis, 2001). The comparison of the proposed equation and already existing well-known equations has been illustrated in the Fig. 19. The proposed GEP equation is in good agreement with the present experimental results as the value of correlation coefficient (R^2) for training and testing was 0.92 and 0.96 respectively. As depicted in Fig. 19, the all the four equations over estimates the scour depth, whereas the proposed GEP equation is in a good agreement with the experimental results. In order to evaluate the prediction accuracy of all the five equations, statistical analysis of the predicted results have been carried out to measure the (Correlation Coefficient) R^2,(Root Mean Square Error) RMSE,(Mean Absolute Error) MAE, (Mean Absolute Percentage Error) MAPE. Results of the statistical analysis are presented in Table 9 and on the bases of these results five equations can be arranged in order of their prediction accuracy such as: proposed GEP equation, Melville and Coleman equation, IRC equation, HEC-18 equation and Larsen and Toch equation.

10. Conclusion

The present results exemplify the influencing parameters for scour depth, which must be considered in order to better comprehend the scour phenomenon, such as pier spacing, flow depth, rate of flow, Froude number, Reynolds number, hydraulic depth, and relative velocity. The scour depth around the front pier in the tandem arrangement was observed to be *21.6%* greater than the single pier and *16%* greater than the scour at the rear pier. As the P_s/d ratio increases, the extent of scour depth increases up to the $P_s/d = 3$, and behaves as a single body. For $Ps/d > 3$, scour depth started reducing and both the piers behaved individually. The scour depth increases with the increase in the F_d and u/u_c. The maximum and minimum Y_s/d was observed at highest F_d *(Fd84 = 1.83)* and lowest number *($F_{d84} = 1.37$)* respectively. Similarly, the

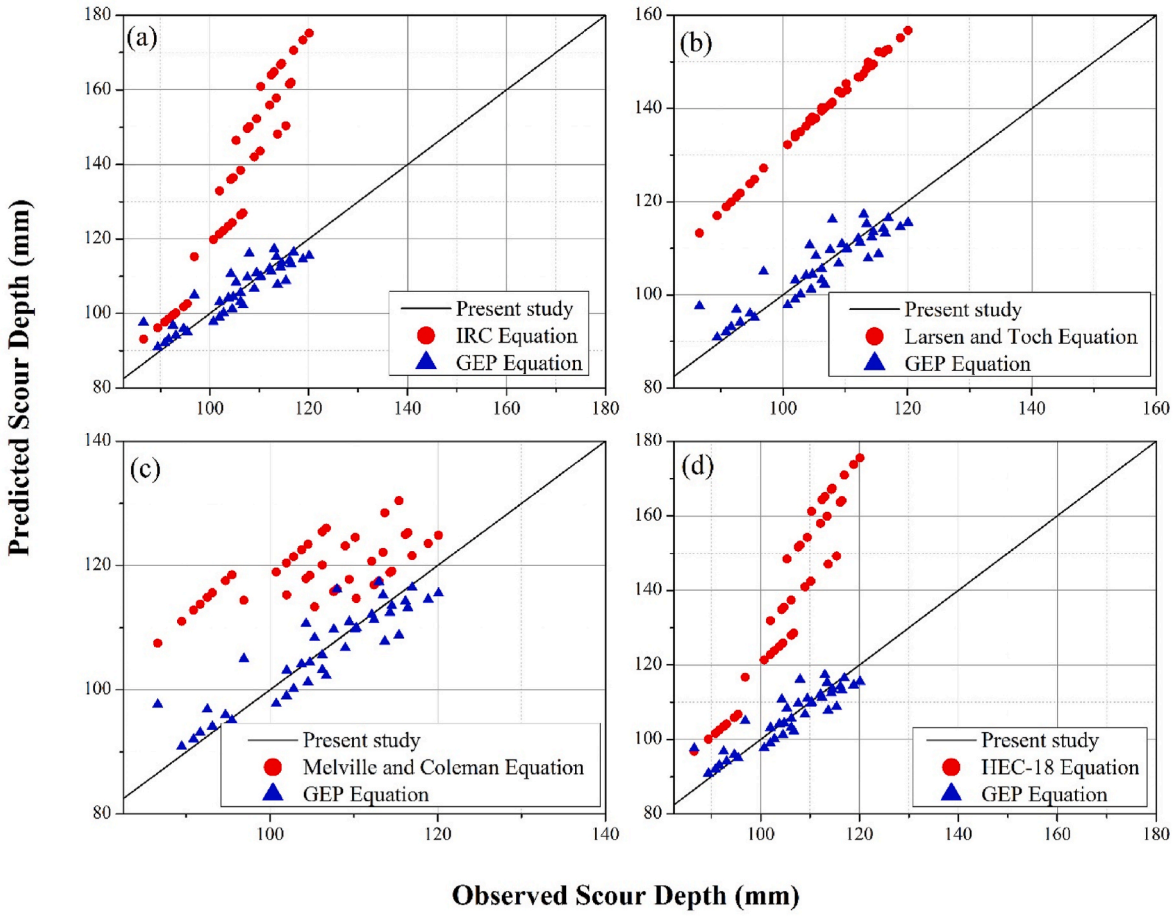


Fig. 19. Comparison of the GEP equation with the existing equations: (a) IRC equation; (b) Larsen and Toch equation; (c) Melville and Coleman equation; and (d) HEC-18 equation.

Table 9
Statistical analysis for the GEP and traditional regression based equations for the present experimental results.

Performance Parameters	GEP	IRC Equation	Larsen and Toch Equation	Melville Equation	HEC-18 Equation
R^2	0.96	0.93	0.99	0.476	0.92
RMSE	3.77	35.14	32.99	15.02	35.92
MAE	2.83	31.06	32.87	13.57	32.41
MAPE	2.7	28.3	31.10	13.4	29.7

maximum and minimum Ys/d was achieved at $u/u_c = 0.98$, and 0.95 respectively.

This study also describes a novel and efficient method for predicting local scour around two piers in tandem positioning utilizing GEP modeling system. The recommended GEP formulations agree perfectly with the experimental results. The proposed equations are so simple that anyone who is not familiar with GEP can use them. The proposed GP formulations results are compared to the existing regression-based equations, and the predictions of GEP outperform. The findings of this study are very promising, and they encourage the use of GEP in complex water engineering problems.

CRediT authorship contribution statement

Geeta Devi: Conceptualization, Writing – review & editing, Visualization, Project administration. **Munendra Kumar:** Conceptualization, Visualization, Supervision, Project administration.

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.

References

Aguirre-Pe, J., Olivero, M.L., Moncada, A.T., 2003. Particle densimetric Froude number for estimating sediment transport. J. Hydraul. Eng. 129, 428–437. https://doi.org/10.1061/(asce)0733-9429(2003)129:6(428).

Ahmed, F., Rajaratnam, N., 1998. Flow around bridge piers. J. Hydraul. Eng. 124, 288–300. https://doi.org/10.1061/(asce)0733-9429(1998)124:3(288).

Albert, Molinas, 2004. Bridge Scour in Non Uniform Sediment Mixtures and in Cohesive Materials : Synthesis Report. Federal Highway Administration Publication FHWA-RD-03.

Amini, A., Melville, B.W., Ali, T.M., Ghazali, A.H., 2011. Clear-water local scour around pile groups in shallow-water flow. J. Hydraul. Eng. 138, 177–185. https://doi.org/10.1061/(asce)hy.1943-7900.0000488.

Amini, A., Melville, B.W., Ali, T.M., Ghazali, A.H., 2012. Clear-water local scour around pile groups in shallow-water flow. J. Hydraul. Eng. 138, 177–185.

Amini, A., Melville, B.W., Ali, T.M., 2014. Local scour at piled bridge piers including an examination of the superposition method. Can. J. Civ. Eng. 41, 461–471. https://doi.org/10.1139/cjce-2011-0389.

Azamathulla, H.M., 2012. Gene-expression programming to predict scour at a bridge abutment. J. Hydroinf. 14, 324–331.

Azamathulla, H.M., Jarrett, R.D., 2013. Use of gene-expression programming to estimate Manning's roughness coefficient for high gradient streams. Water Resour. Manag. 27, 715–729. https://doi.org/10.1007/s11269-012-0211-1.

Azamathulla, H.M., Deo, M.C., Deolalikar, P.B., 2008. Alternative neural networks to estimate the scour below spillways. Adv. Eng. Software 39, 689–698.
Azmathullah, H.M., Deo, M.C., Deolalikar, P.B., 2005. Neural networks for estimation of scour downstream of a ski-jump bucket. J. Hydraul. Eng. 131, 898–908.
n.d Beg, M., 2004. Mutual interference of bridge piers on local scour. In: Proceedings 2nd International Conference on Scour and Erosion (ICSE-2), 14-17, November 2004 singapore.
Beg, M., Beg, S., 2015. Scour hole characteristics of two unequal size bridge piers in tandem arrangement. ISH.J. Hydraul. Eng. 21, 85–96. https://doi.org/10.1080/09715010.2014.963176.
Behzad, A.-A., Abolfazl, A.-K., 2013. Flow field around single and tandem piers. Flow, Turbul. Combust. 90, 471–490. https://doi.org/10.1007/s10494-012-9427-7.
Behzad, A.-A., Beheshti, A.A., 2006. Experimental investigation of clear-water local scour at pile groups. J. Hydraul. Eng.ASCE 132 (10), 1100–1104. https://doi.org/10.1061/(ASCE)0733-9429(2006)132:10(1100).
Bozkuş, Z., Özalp, M.C., Dinçer, A.E., 2018. Effect of pier inclination angle on local scour depth around bridge pier groups. Arabian J. Sci. Eng. 43, 5413–5421. https://doi.org/10.1007/s13369-018-3141-2.
Breusers, H.N.C., Nicollet, G., Shen, H.W., 1977. Local scour around cylindrical piers. J. Hydraul. Res. 15, 211–252. https://doi.org/10.1080/00221687709499645.
Cheng, N.-S., Wei, M., Lu, Y., Chiew, Y.-M., 2020. Evaluation of sediment gradation effects on clear-water pier scour with densimetric Froude number. J. Eng. Mech. 146, 04020133 https://doi.org/10.1061/(asce)em.1943-7889.0001871.
Chiew, Y.-M., 1995. Mechanics of riprap failure at bridge piers. J. Hydraul. Eng. 121, 635–643. https://doi.org/10.1061/(asce)0733-9429(1995)121:9(635).
Congress, I.R., 2016. Standard Specifications and Code of Practice for Road Bridges. Irc.
Das, S., Das, R., Mazumdar, A., 2015. Comparison of local scour characteristics around two eccentric piers of different shapes. Arabian J. Sci. Eng. https://doi.org/10.1007/s13369-015-1817-4.
Devi, G., Kumar, M., 2022a. Countermeasures against local scouring at circular bridge piers using collar and combination of slot and collar. In: Jha, R., Singh, V.P., Singh, V., Roy, L.B., Thendiyath, R. (Eds.), River Hydraulics: Hydraulics, Water Resources and Coastal Engineering, vol. 2. Springer International Publishing, Cham, pp. 289–296. https://doi.org/10.1007/978-3-030-81768-8_24.
Devi, G., Kumar, M., 2022b. Estimation of local scour depth around twin piers using gene expression programming (local scour around twin piers). Water Supply. https://doi.org/10.2166/ws.2022.233 ws2022233.
Dey, S., Raikar, R.V., 2007. Characteristics of horseshoe vortex in developing scour holes at piers. J. Hydraul. Eng. 133, 399–413. https://doi.org/10.1061/(asce)0733-9429(2007)133:4(399).
Dey, S., Bose, S.K., Sastry, G.L.N., 1995. Clear water scour at circular piers: a model. J. Hydraul. Eng. 121, 869–876. https://doi.org/10.1061/(asce)0733-9429(1995)121:12(869).
Elliott, K.R., Baker, C.J., 1985. Effect of pier spacing on scour around bridge piers. J. Hydraul. Eng. 111, 1105–1109. https://doi.org/10.1061/(asce)0733-9429(1985)111:7(1105).
Ettema, R., 1976. Influence of Bed Material Gradation on Local Scour. University of Auckland.
Ferreira, C., 2001. Gene Expression Programming: a New Adaptive Algorithm for Solving Problems arXiv preprint cs/0102027.
Garg, V., Setia, B., Singh, V., Kumar, A., 2021. Scour protection around bridge pier and two-piers-in-tandem arrangement. ISH.J. Hydraul. Eng. 1–13. https://doi.org/10.1080/09715010.2021.1874550, 00.
Ghodsi, H., Najafzadeh, M., Khanjani, M.J., Beheshti, A., 2021. Effects of different geometric parameters of complex bridge piers on maximum scour depth: experimental study. J. Waterw. Port, Coast. Ocean Eng. 147, 4021021.
Hannah University of Canterbury, Department of Civil Engineering., C.R., 1978. Scour at Pile Groups : a Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Engineering at the University of Canterbury, Christchurch, New Zealand. Dept. of Civil Engineering, University of Canterbury, Christchurch, N.Z.
Hamidi, A., Siadatmousavi, S.M., 2018. Numerical simulation of scour and flow field for different arrangements of two piers using SSIIM model. Ain Shams Eng. J. 9, 2415–2426. https://doi.org/10.1016/j.asej.2017.03.012.
Hassan, W.H., Jalal, H.K., 2021. Prediction of the depth of local scouring at a bridge pier using a gene expression programming method. SN Appl. Sci. 3, 1–9. https://doi.org/10.1007/s42452-020-04124-9.
Hassan, W.H., Hussein, H.H., Alshammari, M.H., Jalal, H.K., Rasheed, S.E., 2022. Evaluation of gene expression programming and artificial neural networks in PyTorch for the prediction of local scour depth around a bridge pier. Results in Engineering. https://doi.org/10.1016/j.rineng.2022.100353.
Homaei, F., Najafzadeh, M., 2020. A reliability-based probabilistic evaluation of the wave-induced scour depth around marine structure piles. Ocean Eng. 196, 106818.
Hong, Y., Ph, D., Kan, Y., Ph, D., Zeng, J., Lin, H., Ph, D., 2018. Multiphysics coupling model for computing pier scour upon simulation and experiment. J. Comput. Civ. Eng. 1–15. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000716. © ASCE, ISSN 0887-3801 32.
Hosseini, R., Amini, A., 2015. Scour depth estimation methods around pile groups. KSCE J. Civ. Eng. 19, 2144–2156. https://doi.org/10.1007/s12205-015-0594-7.
Jain, R., Lodhi, A.S., Oliveto, G., Pandey, M., 2021. Influence of cohesion on scour at piers founded in clay-sand-gravel mixtures. J. Irrigat. Drain. Eng. 147, 04021046 https://doi.org/10.1061/(asce)ir.1943-4774.0001616.
Jiang, W., Lin, C., 2021. Scour effects on vertical effective stresses and lateral responses of pile groups in sands. Ocean Eng. 229, 109017 https://doi.org/10.1016/j.oceaneng.2021.109017.
Jiang, W., Lin, C., 2022. Scour effects on lateral responses of pile groups in clays. Comput. Geotech. 149, 104851.
Johnson, P.A., Torrico, E.F., 1994. Scour around wide piers in shallow water. Transport. Res. Rec. 66–70.
Jones, J.S., Sheppard, D.M., 2004. Local scour at complex pier geometries. In: Joint Conference on Water Resource Engineering and Water Resources Planning and Management 2000: Building Partnerships, pp. 1–9. https://doi.org/10.1061/40517(2000)409, 104.
Keshavarzi, A., Shrestha, C.K., Melville, B., Khabbaz, H., Ranjbar-Zahedani, M., Ball, J., 2018. Estimation of maximum scour depths at upstream of front and rear piers for two in-line circular columns. Environ. Fluid Mech. 18, 537–550. https://doi.org/10.1007/s10652-017-9572-6.
Khaple, S., Reddy, P., Roberto, H., 2017. Interference of an upstream pier on local scour at downstream piers. Acta Geophys. https://doi.org/10.1007/s11600-017-0004-2.
Lança, R., Fael, C., Maia, R., Pêgo, J.P., Cardoso, A.H., 2013. Clear-water scour at pile groups. J. Hydraul. Eng. 139, 1089–1098. https://doi.org/10.1061/(asce)hy.1943-7900.0000770.
Laursen, E.M., Arthur, T., 1956. Scour around bridge piers and abutments (1956), Iowa Institute of Hydraulic Research. https://core.ac.uk/download/pdf/76635144.pdf.
Liang, F., Wang, C., Huang, M., Wang, Y., 2017. Experimental observations and evaluations of formulae for local scour at pile groups in steady currents. Mar. Georesour. Geotechnol. 35, 245–255. https://doi.org/10.1080/1064119X.2016.1147510.
Ma, L., Wang, L., Guo, Z., Jiang, H., Gao, Y., 2018. Time development of scour around pile groups in tidal currents. Ocean Eng. 163, 400–418. https://doi.org/10.1016/j.oceaneng.2018.06.035.
Malik, R., Setia, B., 2018. Local scour around closely placed bridge piers. ISH.J. Hydraul. Eng. 1–8. https://doi.org/10.1080/09715010.2018.1559772, 00.
Malik, R., Setia, B., 2020. Interference between pier models and its effects on scour depth. SN Appl. Sci. 2, 1–12. https://doi.org/10.1007/s42452-019-1868-3.
Melville, B.W., Chiew, Y., 1999. Time scale for local scour at bridge piers. J. Hydraul. Eng. 125 (1), 59–65.
Melville, B.W., Coleman, S.E., 2000. Bridge Scour. Water Resources Publication.
Melville, B.W., Raudkivi, A.J., 1977. Flow characteristics in local scour at bridge piers. J. Hydraul. Res. 15, 373–380. https://doi.org/10.1080/00221687709499641.
Memar, S., Zounemat-Kermani, M., Beheshti, A., Rahimpour, M., De Cesare, G., Schleiss, A.J., 2020. Influence of collars on reduction in scour depth at two piers in a tandem configuration. Acta Geophys. 68, 229–242. https://doi.org/10.1007/s11600-019-00393-0.
Mia, M.F., Nago, H., 2003. Design method of Time-Dependent local scour at circular bridge pier. J. Hydraul. Eng. ASCE 129, 420–427. https://doi.org/10.1061/(ASCE)0733-9429(2003)129:6(420).
Mia, F., Nago, H., Asce, F., 2003. Discussion of " Design method of Time- Dependent local scour at circular bridge piers. J. Hydraul. Eng. 129, 1211–1215. https://doi.org/10.1061/:(ASCE)0733–9429(2003)129:6(420) (Giuseppe).
Mohammed Saleh, L.A., Majeed, S.A.A. din, Alnasrawi, F.A., el kadhium, M., 2019. Numerical study of the bridge pier scour using gene expression programming. J. Appl. Water. Eng. Res. 7, 287–294. https://doi.org/10.1080/23249676.2019.1684390.
Muzzammil, M., Gangadhariah, T., 2010. The mean characteristics of horseshoe vortex at a cylindrical pier. J. Hydraul. Res. 41, 285–297. https://doi.org/10.1080/00221680309499973.
Muzzammil, M., Alama, J., Danish, M., 2015. Scour prediction at bridge piers in cohesive bed using Gene Expression Programming. In: International Conference on Water Resources, Coastal and Ocean Engineering: Aquatic Procedia. Elsevier B.V., pp. 789–796. https://doi.org/10.1016/j.aqpro.2015.02.098
Najafzadeh, M., 2015. Neuro-fuzzy GMDH systems based evolutionary algorithms to predict scour pile groups in clear water conditions. Ocean Eng. 99, 85–94.
Najafzadeh, M., Azamathulla, H.M., 2015. Neuro-fuzzy GMDH to predict the scour pile groups due to waves. J. Comput. Civ. Eng. 29 https://doi.org/10.1061/(asce)cp.1943-5487.0000376.
Najafzadeh, M., Oliveto, G., 2021. More reliable predictions of clear-water scour depth at pile groups by robust artificial intelligence techniques while preserving physical consistency. Soft Comput. 25, 5723–5746. https://doi.org/10.1007/s00500-020-05567-3.
Najafzadeh, M., Balf, M.R., Rashedi, E., 2016. Prediction of maximum scour depth around piers with debris accumulation using EPR, MT, and GEP models. J. Hydroinf. 18, 867–884. https://doi.org/10.2166/hydro.2016.212.
Pandey, M., Oliveto, G., Pu, J.H., Sharma, P.K., Ojha, C.S.P., 2020a. Pier scour prediction in non-uniform gravel beds. Water (Switzerland) 12, 13–17. https://doi.org/10.3390/W12061696.
Pandey, M., Zakwan, M., Sharma, P.K., Ahmad, Z., 2020b. Multiple linear regression and genetic algorithm approaches to predict temporal scour depth near circular pier in non-cohesive sediment. ISH.J. Hydraul. Eng. 26, 96–103. https://doi.org/10.1080/09715010.2018.1457455.
Pu, J.H., Wallwork, J.T., Khan, M.A., Pandey, M., Pourshahbaz, H., Satyanaga, A., Hanmaiahgari, P.R., Gough, T., 2021. Flood suspended sediment transport: combined modelling from dilute to hyper-concentrated flow. Water (Switzerland) 13, 1–24. https://doi.org/10.3390/w13030379.
Qi, M., Chiew, Y.-M., Hong, J.-H., 2013. Suction effects on bridge pier scour under clear-water conditions. J. Hydraul. Eng. 139, 621–629. https://doi.org/10.1061/(asce)hy.1943-7900.0000711.
Qi, W.G., Li, Y.X., Xu, K., Gao, F.P., 2019. Physical modelling of local scour at twin piles under combined waves and current. Coast. Eng. 143, 63–75. https://doi.org/10.1016/j.coastaleng.2018.10.009.
Richardson, E.V., Davis, S.R., 2001. Evaluating scour at bridges, National Highway institute (US). https://www.fhwa.dot.gov/engineering/hydraulics/pubs/hec/hec18.pdf.

Setia, B., 2008. Equilibrium scour depth time. In: 3rd IASME/WSEAS Int. Conf. On Water Resources, Hydraulics & Hydrology (WHH '08). University of Cambridge, UK, pp. 114–117. https://www.researchgate.net/publication/267365392.

Sharafati, A., Tafarojnoruz, A., Yaseen, Z.M., 2020. New stochastic modeling strategy on the prediction enhancement of pier scour depth in cohesive bed materials. J. Hydroinf. 22, 457–472. https://doi.org/10.2166/hydro.2020.047.

Shen, H.W., Schneider, V.R., Karaki, S., 1969. Local scour around bridge piers. J. Hydraul. Div. 95, 1919–1940.

Sheppard, D.M., Miller Jr., William, 2006. Live-Bed local pier scour experiments. J. Hydraul. Eng. ASCE 132, 635–642. https://doi.org/10.1061/(ASCE)0733-9429(2006)132:7(635).

Suk, H., Nabi, M., Kimura, I., Shimizu, Y., 2014. Numerical investigation of local scour at two adjacent cylinders. Adv. Water Resour. 70, 131–147. https://doi.org/10.1016/j.advwatres.2014.04.018.

Sumer, B.M., Fredsøe, J., Bundgaard, K., 2005. Global and local scour at pile groups. In: The Fifteenth International Offshore and Polar Engineering Conference. OnePetro.

Tafarojnoruz, A., Gaudio, R., Calomino, F., Marone, S.V., Calabria, U., Cs, R., 2012. Bridge pier scour mitigation under steady and unsteady flow conditions. Acta Geophys. 60, 1076–1097. https://doi.org/10.2478/s11600-012-0040-x.

Tang, H.W., Wang, H., Liang, D.F., Lv, S.Q., Yan, L., 2013. Incipient motion of sediment in the presence of emergent rigid vegetation. J. Hydro.Environ. Res. 7, 202–208. https://doi.org/10.1016/j.jher.2012.11.002.

Tipireddy, R.T.R., Barkdoll, B.D., Asce, F., 2019. Scour reduction by air injection at a cylindrical bridge pier : experimental determination of optimal configuration. J. Hydraul. Eng. https://doi.org/10.1061/(ASCE)HY.1943-7900.0001555. © ASCE, ISSN 0733-9429. 145, 1-4.

Vijayasree, B.A., Eldho, T.I., 2021. A modification to the Indian practice of scour depth prediction around bridge piers. Curr. Sci. 120, 1875–1881. https://doi.org/10.18520/cs/v120/i12/1875-1881.

Wang, H., Tang, H., Liu, Q., Wang, Y., 2016a. Local scouring around twin bridge piers in open-channel flows. J. Hydraul. Eng. 142, 06016008 https://doi.org/10.1061/(asce)hy.1943-7900.0001154.

Wang, H., Tang, H.W., Xiao, J.F., Wang, Y., Jiang, S., 2016b. Clear-water local scouring around three piers in a tandem arrangement. Sci. China Technol. Sci. 59, 888–896. https://doi.org/10.1007/s11431-015-5905-1.

Yanmaz, A.M., Altinbilek, H.D., 1991. Study of time-dependent local scour around bridge piers. J. Hydraul. Eng. ASCE 117, 1247–1268. https://doi.org/10.1061/(ASCE)0733-9429(1991)117:10(1247).

Zhou, K., Duan, J.G., Bombardelli, F.A., 2020. Experimental and theoretical study of local scour around three-pier group. J. Hydraul. Eng. 146, 04020069 https://doi.org/10.1061/(asce)hy.1943-7900.0001794.

This article has been accepted for publication in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2022.3204909

Exploring Bitslicing Architectures for Enabling FHE-assisted Machine Learning

Soumik Sinha, Sayandeep Saha, Manaar Alam, Varun Agarwal, Ayantika Chatterjee, Anoop Mishra, Deepak Khazanchi and Debdeep Mukhopadhyay *Senior Member, IEEE*

***Abstract*—Homomorphic Encryption (HE) is the ultimate tool for performing secure computations even in untrusted environments. Application of HE for Deep Learning (DL) inference is an active area of research, given the fact that DL models are often deployed in untrusted environments (e.g. third-party servers) yet inferring on private data. However, existing HE libraries (*somewhat* (SWHE), *leveled* (LHE) or *fully* homomorphic (FHE)) suffer from extensive computational and memory overhead. Few performance optimized high-speed homomorphic libraries are either suffering from certain approximation issues leading to decryption errors or proven to be insecure according to recent published attacks. In this paper, we propose architectural tricks to achieve performance speedup for encrypted DL inference developed with exact HE schemes without any approximation or decryption error in homomorphic computations. The main idea is to apply quantization and suitable data packing in the form of bitslicing to reduce the costly noise handling operation, Bootstrapping while achieving a functionally correct and highly parallel DL pipeline with a moderate memory footprint. Experimental evaluation on the MNIST dataset shows a significant ($37X$) speedup over the non-bitsliced versions of the same architecture. Low memory bandwidths ($700MB$) of our design pipelines further highlight their promise towards scaling over larger gamut of Edge-AI analytics use cases.**

***Index Terms*—Homomorphic Encryption, Accelerators, Bitslicing, Encrypted Analytics;**

I. Introduction

[1]

Advanced analytics workloads, involving Machine Learning (ML) or Deep Learning (DL), are often deployed in public cloud servers such as AWS, Azure, GCP so that the vast compute infrastructures could be exploited to attain better performance and scalability. Considering the recent vulnerabilities reported in these public servers as published in reports due to [1]–[3], enforcing the security of mission critical ML workloads deployed in cloud servers is of utmost importance. *Encrypted analytics* is a viable solution in this direction with the support of a family of HE scheme.

Most of the existing approaches for encrypted DL are constructed over LHE or SWHE schemes, both of which support homomorphic operations only for limited circuit depth [4]–[7]. Beyond this depth, the ciphertext noise grows to an extent where a successful decryption becomes impossible. While SWHE or LHE schemes are still preferred because of performance benefits, their support for limited circuit depth becomes an issue for complex inference tasks where deeper networks (therefore, deeper circuit depths) are required. FHE addresses this issue with bootstrapping which can reduce the ciphertext noise during homomorphic computation, and, therefore, enables computation for arbitrary circuit depths [8]. Bootstrapping "refreshes" a ciphertext by running the decryption function on it homomorphically, thus reducing the ciphertext noise. However, such power comes at a cost of Bootstrapping, which is both computationally intensive and memory hungry. In [9], authors proposed a FHE bootstrapping mechanism which takes less than 1 second. Later, works due to Chilotti et al. further improved this timing in their work [10], which proposes a Torus polynomial based FHE bootstrapping procedure within 13ms and also compresses the bootstrapping key size from 1 GB to 16MB thus hugely optimizing the memory overhead. We shall revisit this scheme in later sections of this paper to explain its significance in connection to our proposed encrypted ML architecture.

While the choice between LHE/SWHE and FHE can still be debated over based on the performance issue, recent research has shown that FHE is indeed a safer choice from the perspective of security. It has been found that the secret key of certain LHE schemes (such as CKKS [11]) can be extracted with the knowledge of a few decrypted ciphertexts [12], [13]. Such attack can fully compromise the privacy, which is the sole purpose of choosing encrypted analytics. Further, many efficient FHE assisted ML approaches use approximations in the underlying homomorphic scheme to avoid costly bootstrapping. These encrypted ML models only support approximations of nonlinear activations which incorporates approximation errors or support only specific activation functions, like square

Soumik Sinha and Ayantika Chatterjee are with the Advanced Technology Development Center, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal, India 721302.
E-mail: soumiksinha21@gmail.com, ayantika@atdc.iitkgp.ac.in.
Debdeep Mukhopadhyay is with the Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal, India 721302.
E-mail: debdeep@cse.iitkgp.ac.in
Anoop Mishra and Deepak Khazanchi are with the University of Nebraska Omaha.
E-mail: {amishra,khazanchi}@unomaha.edu.
Sayandeep Saha is with Nanyang Technological University, Singapore.
E-mail: sayandeep.saha@ntu.edu.sg
Varun Agarwal is with Delhi Technological University, Delhi, India.
E-mail: agarwalvarun2000@gmail.com.
Manaar Alam is with New York University, Abu Dhabi. E-mail: alam.manaar@gmail.com.

[1]Manuscript received April 07, 2022; revised June 11, 2022; accepted July 05, 2022. This article was presented in the International Conference on Compilers, Architectures, and Synthesis for Embedded Systems (CASES) 2022 and appears as part of the ESWEEK-TCAD special issue.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:52:22 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2022.3204909

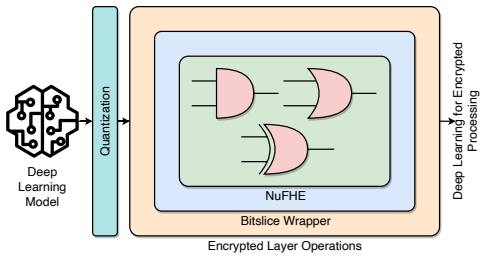


Fig. 1: Overview of the proposed framework

activations [4], [14]. In this situation, FHE schemes with efficient bootstrapping support seems to be a judicious choice for realizing DL inferences.

Noteworthy here is the fact that contrary to traditional algorithm implementations, FHE domain implementations require circuit based realization of algorithms on our traditional unencrypted processors, which is a non-trivial task. This is due to the fact that existing FHE libraries support basic computational gates and homomorphy can only be achieved by realizing algorithms on top of them [10]. Again to support arbitrary noise free FHE computations, usage of costly Bootstrapping operation is mandatory, but that leads to serious performance bottleneck and huge ciphertext size blow-up. Moreover, FHE gates can only support integer domain computations, which is a major drawback to realize any ML model directly in FHE domain. All these limitations motivate us to make an effort in developing a robust FHE encrypted ML inference framework with performance acceleration and without any inference accuracy loss with respect to unencrypted domain. It is to note that, in this era of Machine Learning as a Service (MLaaS), both model parameters and data in ML scenarios can be considered as intellectual property (IP). However, following existing literature and considering the overhead of HE schemes, the threat model of our proposed framework takes data privacy into account and it is developed based on public model private data assumption. Lately, FHE based ML architectures with reduced timing overheads have been proposed by state-of-the-art libraries like Concrete [15]. However, our investigations in this work highlight how Concrete FHE library may lead to erroneous decryption. That is why we build up the proposed encrypted ML inference on mathematically accurate NuFHE [16], which is the extension of FHE library over Torus (supports fastest bitwise homomorphy in literature within few milliseconds) [17]. Our proposed end-to-end encrypted inference pipeline never tweaks the library or does not require any approximation in activation implementation due to the support of basic gate level homomorphy. However, straightforward realization of ML models only with FHE gates incur huge timing as well as memory overhead. As an example, our baseline implementation with only MNIST dataset requires around 80 hours for FHE inference and 6.7 GB memory on average.

To reduce this huge timing and memory overhead, we propose a new FHE ML architecture where we primarily design a wrapper to optimize the overall design with two main architectural tricks: Quantization and Bitslicing. Quantization is an approach widely adopted for reducing the memory footprint of DL model so that they can be easily deployed at resource-constrained embedded devices. Reducing the memory footprint is also essential in the context of FHE. Bitslicing, on the other hand, is a parallelization technique primarily used in cryptography [18], where bitwise logical operations are quite prevalent. The main idea is to pack the operands of several bitwise operations in the same register, so that they can be processed in parallel with one processor instruction. In FHE context, bitslicing is mainly adapted as tensor packing technique to reduce the number of Bootstrapping operation. This in turn can boost the performance of both Conv2D layers and Dense operators. Further, the prerequisite of FHE based implementation setting is discretized tensors as opposed to real valued ones and we achieve this using Quantization supported inference. Our design ensures that bitsliced accelerator with FHE operations work on integer ciphertext operands. With all these modifications, following are the major contributions of this paper:

- First, we try to highlight some limitations of state-of-the-art Concrete library which incorporates decryption errors for real valued tensors.
- Further, FHE computations are only possible on integer values and Bootstrapping operations on 32-bit tensors could worsen the overall performance. To mitigate these issues, we adopt quantized inference on $INT8$ or 8-bit Integer tensors. Without accuracy degradation, $INT8$ quantization empowers us in performance boosting.
- Next, we revisit the building blocks of CNN or DNN architectures and optimize them in the FHE context. Finally, we design a memory efficient ML accelerator with architectural tricks like packing and software bitslicing. We evaluate our techniques on both CNNs as well as DNNs and demonstrate the speedup potential of around 37 times.
- We perform an in-depth memory profiling and try to benchmark our inference accelerator with other state-of-the-art approaches reported in [4], [5], [19] and [7]. These reported works not only incorporate approximations, most of them also incur huge memory requirement of around 2-3 GB for evaluating the inferences. Works like [[5], [7]] are low latency pipelines but they are certainly memory intensive (consuming atleast few GB's for inference on moderate architectures). Our bitslicing based design paradigm, however optimizes significantly on memory footprint and are capable of CNN inference within tight memory budgets of around $700MB$.

With this motivation, our paper is organised as follows: Introduction section (Section I) highlights the security vulnerabilities of public cloud servers and explains the motivations to adopt HE for encrypted ML architecture design. Section II highlights the background and limitations of existing works. Section III explains few preliminary concepts with important terminologies. Section IV describes the design details of proposed ML architecture for encrypted Inference accelerated by

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:52:22 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2022.3204909

TABLE I: HE Schemes Taxonomy

HE Scheme	Examples	Open-source Libraries
Partial(PHE)	RSA,EL-Gamal [20], [21]	OpenSSL
Somewhat(SHE)/Leveled(LHE)	BGV [11], [22]	HELib, SEAL Palisade
Fully/FHE	Chilotti's TFHE [10]	NuFHE, Concrete

TABLE II: Concrete Leveled Computation Errors

Plaintext	Level_1 Error	Level_2 Error	Level_3 Error	Level_4 Error
8.2+5.6+1.2+2.2+5.0	0.032	0.028	0.026	0.0321
4.2+5.6+1.2+2.2+5.0	0.00031	0.01022	0.02004	0.0503
10+1.1+0+2.2+5.0	0.022	0.0292	0.0264	0.0329
0+2.2+3.3+4.4+5.5	0.0133	0.0239	0.0264	0.0329
1.0+2.0+3.0+4.0+5.0	0.0097	0.0154	0.0225	0.036

TABLE III: CONCRETE FHE and LHE Errors

Plaintext	Target Function=x^3	FHE Error	LHE Error
-2.0	-8	3.14	.004809
-4.0	-64	5.3174	.00236
7.0	343	3.154	.00772
5	125	7.94	.004809

architectural optimizations like Quantization, Bitslicing etc. This section also includes the optimized design details for FHE counterparts of CNN building blocks. Finally Section V mentions the experimental details and results comparing with state-of-the-art encrypted ML frameworks followed by conclusion in section VI.

II. Background and Limitations of Related Works

In this section, we highlight few important encrypted ML architectures reported in recent literature based on the existing HE libraries as enlisted in Table I. Among these homomorphic ML frameworks, the encrypted networks mentioned in [4] and [23] suffer from liberty of selecting the activation functions. These encrypted networks can only incorporate square activations due to the choice of limited homomorphic schemes. Further, efforts like [24] can only work with approximated layers with low degree polynomials. The work in [7] reports only estimated results for deeper networks and only provides implementation results for moderate sized networks that can be evaluated with predefined multiplicative depth of a leveled HE scheme. Again, the reported FHE ML inference in [14] approximates the existing FHE scheme on Torus [17] to improve performance. However, the scheme suffers from the limitations of the signed activation functions, which sometimes leads to erroneous predictions.

All these aforementioned efforts are based on either some leveled or arithmetically approximated encryption libraries, which incorporate some inherent limitations. Next, we will try to vividly highlight the major bottlenecks of different classes of encrypted ML frameworks which dominate the literature.

1) Binarized or Discretized Inference based encrypted ML approaches: The promise of neural network inference with binarized or discretized tensors have lately caught attention of researchers due to its performance efficiency. Most relevant approaches report the acceleration potential of quantized inference with bit-wise logical operations. As described in [25], this set of models works with Binarized tensors (all tensors represented as $Power\text{-}of\text{-}2$) and hence all multiplications on those tensors could be achieved with logical bit shift operators. This property makes this technique very effective with inference workloads in FHE domain, as in binarized inference the costly FHE multiplications can be replaced with bit shifts thereby saving huge bootstrapping complexity as reported in [7]. Borrowing from the concept of binarized neural networks, the effect of discretized neural networks has further been explored in Bourse's work [14]. While the former methods largely resort to multiplication less inference networks, this method performs some thresholding of activation tensors apriori so that the approximation function computation will come down to a matter of just changing sign bits. Noteworthy here, that all these quantized weights or even predefined thresholds of activation units are applicable to supervised learning paradigms only where they are trained before inference during training phase. As a caveat, this logarithmic quantization trick might not work in such a straightforward manner for other ML use cases which employs linear or ridge regression or some other unsupervised learning algorithm.
2) Approximation schemes for CNN Inference: To avoid costly bootstrapping in terms of memory blowup and performance, many HE assisted ML approaches avoid FHE and opt for approximated leveled HE as underlying scheme. Due to the absence of bootstrapping operation, non-linear activation functions need to be replaced with low degree polynomial functions to be evaluated by leveled HE libraries. For example, [6] is such an approach which approximates non-linear activations with arithmetic polynomials. Inference frameworks proposed in [4] and [24] build up on same ideas and also produce good accuracy. However, scheme proposed in [11] is one of the important approximated library which is getting used to develop the basic building blocks of recent encrypted CNN architectures. Recent attack [12] highlights the vulnerabilities of this underlying HE scheme and thus encrypted ML models designed based on this scheme cannot be considered fully secure anymore.
3) LHE or SWHE based encrypted ML inference: LHE or SWHE only allow limited number of homomorphic computations (limited multiplications or additions). Simpler neural net architecture might run smoothly with underlying LHE libraries as reported in literature [4], [24]. However, with the growth of multiplicative depth of neural networks, ciphertext noise amplifies to an extent when successful decryption is no longer possible. For complex real-life use cases, inference on deeper nets may become erroneous (and training is infeasible). That makes the acceptance of LHE and SWHE schemes very limited for deep neural networks. To resolve this limitation of SWHE or LHE

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:52:22 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully
content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2022.3204909

schemes, usage of FHE have been proposed. Recently, works like [14], [26], [27] demonstrate efforts in building neural net inference frameworks which runs on FHE encrypted ciphertexts with bootstrapping.

4) Concrete based TFHE encrypted ML inference: Zama's Concrete library is an optimized library for FHE ML inference [15]. However, our experimental observations point out that homomorphic operations supported by this library incurs decryption errors computed on integer plaintexts. To elucidate on this, we perform some cubic functions on ciphetexts and evaluate few leveled computations. Table-III and II highlight errors that Concrete [15] encounters during plaintext decryptions during Bootstrappable FHE or LHE computations (Level_i Error in the table indicates error occurred after i-th addition). These errors can amplify as the level grows and the approximations performed by Concrete can lead to misleading inference in ML processing. Concern of significant accuracy drop (around 80%) is already reported in literature [26] for deeper networks and that questions the consistency of Concrete library for encrypted ML architecture design.

In this context, our objective is to keep the underlying FHE encryption library intact and provide architectural optimizations to achieve overall performance improvement. To summarize, in this work we try to address few major challenges ahead:

1) Floating point inputs or weights vectors cannot be processed by FHE schemes, which will be handled by suitable quantization without any accuracy loss.
2) Design of scalable building blocks for FHE inference pipeline for example Conv2D, Dense, MaxPool and AvgPool will be while ensuring low bootstrapping complexity of FHE operations in the process.
3) Deep Neural Networks (DNNs) involve computation of activation functions (like tanh, sigmoid, ReLU). It is a generic challenge for any limited HE scheme to realize these functions. Our proposed architecture will be capable enough to support all kind of activation without any approximation error.

III. Preliminaries

In this section, we elaborate some HE primitives important for further discussion of this paper. HE is an elegant mathematical tool which allows computations on the ciphertext without any explicit decryption. Most HE schemes support at least one of the two fundamental computations either homomorphic addition (HADD) and homomorphic multiplication (HMUL). For example, the classical RSA cipher is a multiplicative HE, which supports multiplication operation in encrypted domain. According to the power of encrypted computation, HE schemes can be classified broadly into three categories: Partial Homomorphic Encryption (PHE) allows only one homomorphic operation either HADD or HMUL; Somewhat Homomorphic Encryption (SWHE) allows limited number of one operation, generally HMUL; Leveled Homomorphic Encryption (LHE) allows limited number of homomorphic operations and finally Fully Homomorphic Encryption (FHE) allows arbitrary number of computations. FHE schemes, in general contain elementary homomorphic gate operations like AND, OR, NAND etc followed by Bootstrapping which is a crucial operation for denoising in FHE circuits. To understand the concept of noise in homomorphic circuits, consider the following example [28]:

Considering a homomorphic encryption scheme with odd number p as shared secret key:

- To encrypt a bit m, a random large q and small r are chosen and output ciphertext is computed as $c = pq + 2r + m$. Ciphertext is close to a multiple of p and m = LSB of distance to nearest multiple of p.

- The decryption is performed as: $m = (c \bmod p) \bmod 2$.

Now, we explain why the above mentioned scheme is considered to be homomorphic. Considering two ciphertexts $c_1 = q_1 p + 2r_1 + m_1$ and $c_2 = q_2 p + 2r_2 + m_2$, the addition and multiplication operations are defined with the following equations. Addition of the ciphertexts can be defined as:

$$\begin{aligned} c_1 + c_2 &= (q_1 + q_2)p + \underbrace{2(r_1 + r_2) + (m_1 + m_2)}_{\text{Distance to nearest multiple of p}} \\ c_1 + c_2 \bmod p &= \underbrace{2(r_1 + r_2)}_{\text{error-term}} + (m_1 + m_2) \end{aligned}$$

Multiplication of the ciphertexts can be defined as:

$$\begin{aligned} c_1 \cdot c_2 &= (c_1 q_2 + q_1 c_2 - q_1 q_2)p + 2(2r_1 r_2 + r_1 m_2 + \\ &\quad m_1 r_2) + m_1 m_2 \\ c_1 \cdot c_2 \bmod p &= \underbrace{2(2r_1 r_2 + \ldots)}_{\text{error-term}} + m_1 \cdot m_2 \end{aligned}$$

Homomorphic property retains till the error-term is within a certain limit. However, the error term (or noise) grows faster for multiplication compared to addition. Hence, the scheme is somewhat homomorphic. Gentry's work [28] has shown the first noise handling technique termed as Bootstrapping and explained the first plausible construction of fully homomorphic encryption scheme.

Bootstrapping a ciphertext indicates homomorphically decrypting it, using a homomorphic encryption of its own secret key. In the end, we get an encryption of the same plaintext, but with reduced noise level. In practice, this operation is the main support to retain homomorphy for arbitrary computations but it is again the main reason for serious performance bottleneck and memory hungry nature of FHE. Hence, FHE literature is still evolving to improve the Bootstrapping technique. Among other existing schemes, Torus based FHE scheme proposed on the concept of Learning with Errors (LWE) paradigms [29], [30] offers fastest gate bootstrapping within few milliseconds.

Our proposed HE ML architecture is implemented based on python library NuFHE [16], which is an extension of that Torus FHE (TFHE) scheme [10]. Hence, we restrict our HE related discussion in context of TFHE. More formally, any FHE scheme can be defined as a tuple of four algorithms : (Gen, Enc, Dec and Eval). Considering security parameter λ and two public parameters N and k, the algorithms related to TFHE can be defined as follows:

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:52:22 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2022.3204909

- Gen (1^λ, N, k): This keygeneration algorithm generates secret key Sk and evaluation key Ek, which will be used in subsequent algorithms.
- Enc(m, Sk): This encryption algorithm encrypts message $m \in \mathcal{M}$, where message space $\mathcal{M} = \mathbb{T}[X]/(X^N + 1)$ (N sized Torus polynomial defined over Torus $\mathbb{T}$). Since, the secret key Sk consists of k number of N-sized polynomials, the ciphertext Ct is computed as (A, B), where A is a random public parameter and $B = \sum_{i=1}^{k} A[i] \cdot Sk[i] + m + e$. Parameter e is Ring-LWE (RLWE) noise polynomial, where each e_j is chosen from Gaussian distribution.
- Eval (Ct, F, Ek): This homomorphic evaluation step evaluates arbitrary function F on RLWE ciphertext Ct by intermediate LWE ciphertext generation. In FHE mode, homomorphic evaluation incorporates circuit bootstrapping procedure, that takes a LWE ciphertext as input, reduces its noise, and converts it back to an GSW ciphertext suitable for subsequent packed operations. Later on the final LWE output from evaluation technique is converted to final RLWE ciphertext.
- Dec (Ct, Sk): Final decryption step involves computation of $\phi = B - \sum_{i=1}^{k} A[i] \cdot Sk[i]$ to get $m + e$. Finally, each of the N coefficients of ϕ are rounded up to return the coefficients of m.

Consequently, a couple of Torus based libraries have been publicly released by Chilotti et al. of which TFHE([17]), Zama's Concrete ([31]) and NuFHE [16] are widely being used to this date.

A. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) proposed by [32] have evolved into widely popular classification algorithms for advanced analytics tasks, credited to their phenomenal success in the image or video classification domain. A CNN is a multi-layered neural network with three following layers:

1) *Convolution:* The convolution operation extract features from an image through the dot product of image pixels and a custom image kernel or image filter. A $H \times W$ image is typically convoluted with $M \times M$ kernel to get $(H - M + 1) \times (W - M + 1)$ output. In convolution, operations like stride and pooling are integrated to preserve the dimensions of the image. Feature maps are generated in convolution operation, which provides insights into the features extracted from the convolution. For 2D convolution operation, each feature map is a 2D matrix that contains features as data vectors that can be visualized on the image space.
2) *Activation Function:* Activation functions are used to transform the weighted sum of the inputs with bias into an output. It is used to provide non-linearity to the output. ReLU, softmax, and sigmoid are a few most used activation functions in CNN.
3) *Pooling:* Pooling is a downsampling mechanism for the image and is usually applied after the convolution operation. It reduces the computation load. MaxPool and Average Pool (AvgPool) are the most commonly used pooling mechanism where maximum and average values are taken from the matrix produced by convolution layer.

TABLE IV: Accuracy with/without Quantization

Dataset	Test/Train Split	Accuracy ($FP32$)	Accuracy ($INT8$)
MNIST	48000/12000	96.43 %	96.33 %
Mendeley	30488/7622	96.82 %	97.24 %

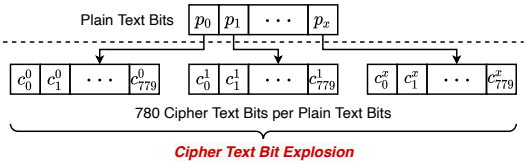


Fig. 2: Blowup in ciphertext size

B. Bitslicing in Software

Originally proposed by Eli Biham [18], for accelerating the computation of DES [33] in software, bitslicing is a parallelization technique widely utilized for cryptographic implementations. The main idea is to utilize the (bitwise) logical instructions (AND, XOR, NOT, OR) present in modern processors in a SIMD fashion. The entire computation is first represented as a circuit consisting bitwise logical operations. Next, a number of parallelizable bit operations are packed together in registers and processed with logical instructions. For example, if a processor architecture contains n-bit registers, and there are n bitwise logical-AND operations in the computation that can happen in parallel, then the operands are packed in two registers and processed with a single bitwise AND instruction. This trick computes AND for all n-bits in parallel, which results in significant speedup. For state-of-the-art block ciphers, which can always be represented as a sequence of bitwise operations, such bitslicing can enable parallel processing of n plaintexts, where n is the width of the registers in the processor.

IV. Quantized Inference with Bitslicing

The goal of this research is to address the challenges of realizing large and complex neural networks in FHE domain with reasonable timing and memory footprints. This is achieved through two basic steps – quantization and bitslicing. We note that quantization is a well-known model compression technique in DL research. However, we found that it is also an essential step for implementing FHE-encrypted DL. The impact of quantization is explained in the next subsection. We also present some basic notations in Table. VI, which we shall use throught the following sections.

A. Quantized Inference of CNNs

In simple words, model quantization aims to reduce the precision of the trained weights (and inputs) in a model, so that they can fit within a reduced bit width. Mathematically, quantization maps all elements in the range $[R_1, R_2]$ to the range $[-(2^{(B-1)}), (2^{(B-1)} - 1)]$, where B is the bit-precision

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:52:22 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2022.3204909

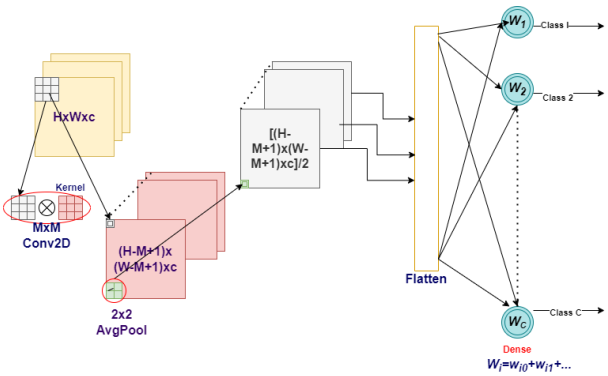


Fig. 3: *CNN Architecture Block Diagram*- this model architecture is has been adopted for MNIST $(W = 28, H = 28, M = 3, c = 1)$ as well as Mendeley Bridge $(W = 224, H = 224, M = 3, c = 3)$ datasets which are then mapped to bitsliced simulations of FHE circuits

TABLE V: DL Architectures (Mendeley and MNIST)

Layer	Output-shape (Mendeley)	Trainable Parameters (Mendeley)	Output-shape (MNIST)	Trainable Parameters (MNIST)
input_layer	(224,224,3)	3	(28,28)	3
reshape	(224,224,3)	1	(28,28,1)	1
Conv2d	(222,222,3)	27	(26,26,1)	15
Max_pooling2d	(111,111,3)	1	(13,13,1)	1
Flatten	(36963)	1	(169)	1
Dense	(2)	73926	(10)	1705

after quantization. For example, if 32-bit floating point weights can be compressed to 8-bit weights it leads to a significant reduction in the memory footprint of a model.

In the context of FHE, however, model compression has several other benefits. The memory blowup for FHE is primarily attributed to the ciphertext size. For Boolean FHE operations, each plaintext bit is mapped to a multi-bit ciphertext. High precision (i.e. larger bit width) operands can, therefore, result in excessive memory overhead. Moreover, larger bit-width also results in higher number of ciphertexts which eventually increases the timing overhead. For example, the bootstrapping time increases significantly while processing on 32-bit floating point $(FP32)$ representations, which is a common weight representation in many DL libraries. Significant improvement due to quantization is, therefore, evident.

However, an obvious impact of quantization is the degradation in inference accuracy. Fortunately, such accuracy loss can be bounded to a reasonable limit, thanks to the efficient quantization routines implemented in modern DL pipelines such as TFLite [34]. In general, at most 2% of accuracy loss is observed due to a quantization from $FP32$ to 8-bit integers $(INT8)$. Even without FHE, this provides a speedup of 2 times over the $FP32$ counterparts [35].

Throughout this paper, we utilize two representative datasets (MNIST [36] and Mendeley [37]) for validating our ideas. The MNIST dataset comprises $60,000$ (28×28) images, all representing digits between $(0 \ldots 9)$. The Mendeley dataset

TABLE VI: Notations and Terminologies

Symbol	Description
p_{ij}	ij^{th} pixel of image matrix
w_{ij}	ij^{th} weight value
H	Height of the Image Matrix
W	Width of the Image Matrix
M	Dimension of $(M \times M)$ Convolution Filter
N	Packing width for bitslicing
c	Number of channels

is dedicated to the task of developing a binary classifier for concrete building images labeled as cracks and no-cracks. Each label class has $20K$ images each having dimensions of $224 \times 224 \times 3$.

We first observe the performance of the models corresponding to these datasets after quantization. It is evident that not every model is supposed to give the same accuracy after quantization. We figured out CNN architectures for both datasets which retain the accuracy even after quantization. In fact, for the Mendeley dataset we observed slightly higher accuracy for the $INT8$ quantized version than the baseline $FP32$. While such increase in accuracy is rare, we note that it is not impossible. For example, such increase in accuracy has also been observed in [34], which also depends on the range of the quantization. Our reference architecture (see. Fig. 3 and Table. V) remains the same for both datasets except changes in few hyper-parameters. More precisely, for MNIST we utilize $(W = 28, H = 28\ M = 3, c = 1)$, where $W \times H$ is the dimension of the input image, and $M \times M$ is the dimension of the CNN filters. c denotes the number of channels in the image. The model also uses AvgPool as well as flattening layers. For the Mendeley dataset, the parameters are $(W = 224, H = 224$ $M = 3, c = 3)$.

The inference accuracy of the models after quantization is presented in Table. IV. It can be observed that the degradation in accuracy for MNIST is quite reasonable. On the contrary, for Mendeley we observe a slight increase in accuracy. While we believe that such observation is specific to this model instance, this has been observed previously for multiple use cases [38], [39]. Overall we achieve reasonable accuracy, while compressing the model significantly making it amenable to FHE conversion.

B. Bitslicing

Quantization is indeed beneficial in terms of timing and memory. However, it is not sufficient to accelerate the FHE versions of a DL model. A straightforward FHE implementation of this model (for MNIST) with NuFHE library takes almost 80 hours and 6.7 GB (on average) of memory to make a single inference. NuFHE generates a 780-bit ciphertext corresponding to each bit of computation and leverages the FHE gates to process them. The blowup in ciphertext size is illustrated in Fig. 2. The memory and computational overhead is evident from this blowup. A crucial question is can these performance figures be improved?

While one might have several attractive results for LHE and SWHE libraries (or for FHE libraries like Concrete with some

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:52:22 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2022.3204909

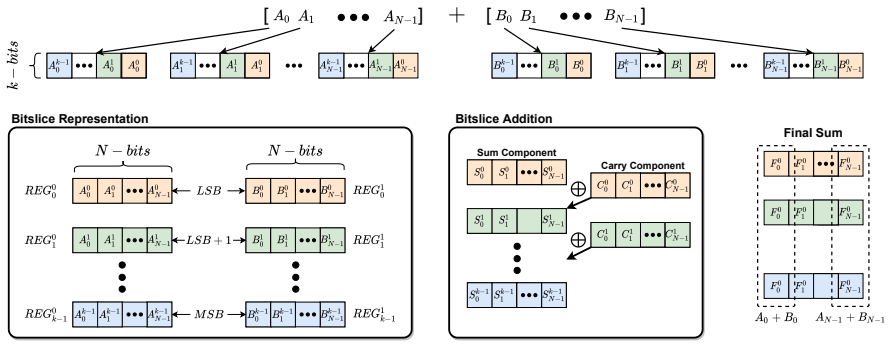


Fig. 4: Bitsliced Full Adder

output errors), none of these is an ultimate solution as pointed out in the introduction. Therefore, we focus on improving the timing and memory overhead for NuFHE library which contains bootstrapped Boolean gates. In this work, Conv2D, MaxPool, AvgPool and Dense layers are implemented using gates from NuFHE. As we explain throughout this section, there are several points of performance optimizations while implementing these operations which result in significant performance boost. The core component behind these performance boost is a bitsliced adder module, which we describe in the next subsection. At this point it is important to note that none of our implementations including the bitsliced adder need any augmentation to be made in the NuFHE library. Rather, we build a software wrapper over the library to implement the layer operations, which is later utilized to construct the DL models.

1) Bitsliced Full Adder (BFA): The main idea behind our bitsliced full Adder module is presented in Fig. 4. It performs addition between two arrays of integers in a bitsliced manner. Referring to Fig. 4, the two lists to be added are denoted as $A_{lst} = [A_0, A_1, \cdots A_{N-1}]$ and $B_{lst} = [B_0, B_1, \cdots B_{N-1}]$. The output is $Sum = [(A_0 + B_0), (A_1 + B_1), \cdots (A_n + B_n)]$. Without loss of generality let us assume that each A_i (resp. B_i) is represented as $A_i = \langle A_i^{k-1}, \cdots A_i^1, A_i^0 \rangle$ with k bits. A straightforward addition of A_{lst} and B_{lst}, therefore, requires N invocations of a k-bit full adder. Each of these k-bit adder invocation requires $4k$-XORs, $2k$ AND operations (assuming a ripple carry adder). Since each gate operation in NuFHE is bootstrapped, total $N \times 6k$ bootstrapping operations are performed. Even assuming some parallelization (especially for the k XORs computing the sum of two k-bit numbers), the number of bootstappings remain high, as the carry chain cannot be parallelized easily. One of the main optimization challenges in this regard is to reduce the number of bootstrapping. In NuFHE context it implies a reduction in the number of gates which is difficult for a small circuit like an adder. Instead, we exploit the fact that NuFHE gates can process multiple ciphertexts (corresponding to different bits) when they are packed together. We exploit this feature in an SIMD fashion.

In order to improve the efficiency we introduce the concept of *bitsliced packing* $\mathsf{PACK}(j)$ of ciphertexts. More precisely, we utilize k arrays $(REG_0, REG_1, \cdots REG_{k-1})$, each of size N. The REG_j stores the jth least significant (LSB) bit of the integers $\{A_i\}_{i=0}^{N-1}$. The $\mathsf{PACK}(j)$ function is illustrated in Algorithm. 1 and Fig. 4.

Algorithm 1 $\mathsf{PACK}(j)$

Input: $A_1 \ldots A_N$, REG_j
Output: REG_i
1: **for** $t \leftarrow 0$ to $N-1$ **do**
  $REG_j[t] \leftarrow A_t^j$
2: **end for**
3: **return** REG_j

We further demonstrate the workflow of packing algorithm with a short example, where we pack 4 integer operands into 4 copies of 4-bit arrays. Suppose, the integers are $A_{lst} = [1, 3, 5, 6]$ whose binary representations are given by 0001 ,0011 ,0111 ,0110. The packing registers are $REG_0, REG_1, REG_2, REG_3$. In this example, $N = 4$, so as per our notation, the operands are $A_0, A_1, \cdots A_3$ where $A_0 = [A_0^3, A_0^2, A_0^1, A_0^0] = [0001]$. Similarly, $A_1 = [A_1^3, A_1^2, A_1^1, A_1^0] = [0011]$, $A_2 = [A_2^3, A_2^2, A_2^1, A_2^0] = [0111]$, $A_3 = [A_3^3, A_3^2, A_3^1, A_3^0] = [0110]$. After the execution of $\mathsf{PACK}(j)$ for $0 \leq j < 4$, REG_0, REG_1, REG_2, and REG_3 contains [1110], [0111], [0011] and [0000], respectively.

Next, we show how to utilize this bitsliced packing to implement a full-adder module for adding two arrays.

The Adder Circuit: Let us consider two lists A_{lst} and B_{lst} packed in two sets of arrays REG_j^0 and REG_j^1, where $0 \leq j < k$ and $|REG_j^h| = N$ $(h = \{0, 1\})$. Referring to Fig. 4, an $N - bit$ bitwise XOR can be deployed for each (REG_j^0, REG_j^1) pair to calculate the sum component of addition. Similarly, the carry logic can be implemented between these two registers. It is worth noting that for REG_j^h's after the LSB position, there is a carry propagation to enable the ripple carry (see Fig. 4) which further needs N-bit XORs. After the carry propagation completes, the final sum is also stored within one of the REG_j^hs.

Let us now analyze the timing and memory overhead of this bitsliced adder compared to the naïve approach of invoking N, k-bit adders. The number of gate operations essentially remains the same. However, due to the bitsliced packing, each bootstrapped gate works on N bits in BFA. In contrast, the carry

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:52:22 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2022.3204909

chain of naïve k-bit adders work on single bit at a time. Consequently, number of required bootstrappings increase N-fold for the naïve case, compared to the BFA case. This is the sole reason behind the timing improvement that we observe in our implementations. Reduction of bootstrapping also significantly improves the memory overhead (reduces it from 5.2 GB to 280 MB). As already pointed out in the preliminaries, bootstrapping involves several complex operations such as key switching and blind rotate which are highly memory intensive. While multi-threading may improve the timing overhead of bootstrapping to some extent, it is not so effective in the context of memory as each thread consumes around 200 MB, which can quickly blow up with N. Overall, the BFA adder is more efficient for adding two arrays compared to naïve addition. It only requires a software wrapper over the NuFHE library without incurring any significant coding complexity.

2) Bitslicing in DL: We next implement different layer operations of DL using the BFA as a core component. Referring to the generic architecture in Fig. 3, here we implement 3 major operations namely Conv2D, MaxPool2d and Dense layer. These operations are also representative to most of the existing DL layer operations.

Encrypted Conv2D/Dense Operators: The core component of Conv2D or Dense operations can be represented as vector-matrix products. Most of the cases it takes the form $y = \sum_{i=1}^{n} w_i x_i$, where w_is are the weights, and x_is are the inputs or intermediate values. We note that for encrypted inference, the weights are in plaintext and the x_is are encrypted. Therefore, in FHE, such multiplications can be implemented as scalar products followed by an addition. The scalar products, on the other hand, can be implemented using repeated additions. A natural way of implementing these repeated additions is to use a loop with the w_i in loop counter.

We begin our discussion in the regard with the Conv2D operation. Let us assume that a convolution filter of dimension $M \times M$ works on an image with dimension $W \times H$. For the sake of illustration, here we take $M = 3$. In Conv2D, the filter moves over an image in steps called strides (as shown in Fig. 5), calculating the convolution. Referring to Fig. 5, each convolution operation can be represented as $\sum_{i=t_1}^{(t_1+2)} \sum_{j=t_2}^{(t_2+2)} w_{(i \bmod 3)(j \bmod 3)} p_{ij}$, where w_{ij}s represent the weights, p_{ij}s represent the inputs, and (t_1, t_2) represents the row, column index of a stride. Now considering two consecutive row-wise strides, one may observe that the convolution operations contains several product terms having the same multiplicand w_{ij}. For example, in Fig. 5 one may observe that $P_{w_{11}} = (p_{11}, p_{12}, \cdots p_{(W-2)H-2})$ gets multiplied with the same weight w_{11}. Since the scalar product is nothing but consecutive addition, therefore, the scalar product $\langle w_{11}, P_{w_{11}} \rangle$ can be represented as repeated addition of the integer array/vector $P_{w_{11}}$. This repeated addition is implemented by invoking the BFA unit $(w_{11} - 1)$ times.

Once the product terms of convolution are computed, the next step is to perform the addition of these product terms. In order to perform this step, one needs to unpack the bitsliced products. This is a straightforward step and an inverse of the $\mathsf{PACK}(j)$ operation. Interestingly, the unpacking also results in several arrays of integers to be added together. The BFA unit is invoked once again to perform this operation, which eventually returns the convolution output.

Let us now explain the computational complexities due to this bitsliced Conv2D operation. One should note that Conv2D is one of the most costly operations for FHE as it involves several scalar multiplications and additions. The packing width ($|P_{w_{ij}}|$) here determines how much parallelization can be obtained. Considering row-wise movement of the convolution filter, there can be total $(W - M + 1)$ strides which can be parallelized. Therefore, we have $|P_{w_{ij}}| = (W - M + 1)$. On the other hand, for a convolution filter of dimension $M \times M$, there can be total M^2 weight values. As a result, during the addition phase of the product terms, the width of each array/vector will be of M^2, and there will be $(W - M + 1)$ such arrays/vectors. For MNIST images we have, $(W - M + 1) = (28 - 3 + 1) = 26$ and $M^2 = 9$. For Mendelay images, however, we face an implementation obstacle. The NuFHE allows packing of at most 32 ciphertexts, which limits the maximum size of $P_{w_{ij}}$ to 32. For 224×224 images in Mendeley, we therefore need $224/32 = 7$ independent packings to be implemented for a complete row-wise convolution. The overheads for the convolution can be calculated in a similar manner.

The implementation of the encrypted dense layer is similar to the Conv2D layer, in principle. In this case, we figure out (encrypted) inputs having a common weight multiplier. This is a straightforward preprocessing as the weights are in plaintext. Once a set of ciphertexts with a common multiplier is identified, we apply the same bitsliced processing approach as explained for Conv2D to process them in parallel. One critical point here is that unlike the Conv2D case, the width of $P_{w_{ij}}$s for Dense layer will not be equal. Also the unpack step will require some extra effort to properly arrange the product terms for addition. However, handling the first issue is easy, as our approach does not depend upon the length of $P_{w_{ij}}$s. To rearrange the product terms we maintain a dictionary storing the actual location of product terms in the computation. The rearranging after unpack can be performed in constant time only by referring to this dictionary. Therefore, we can obtain similar parallelization as for Conv2D even in the case of Dense.

AvgPool2D and MaxPool2D: The objective of AvgPool2D is to calculate the average of Conv2D outputs. Averaging requires addition followed by division operations. It is worth mentioning that the addition here is not bitsliced. This is because the addition step in AvgPool2D only adds a few integers for which the overall overhead is not very significant. However, the divisions are costly in FHE. The straightforward way of performing division as repeated subtraction has to be ruled out here because results can be in floating points which is not inherently supported by FHE schemes. Moreover, the loop terminating criterion for this repeated subtraction is in encrypted domain, which makes it even more complex to process in FHE.

The first step towards avoiding the costly divisions is to limit ourselves to divisions by 2^l, which (for integers) can be

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:52:22 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2022.3204909

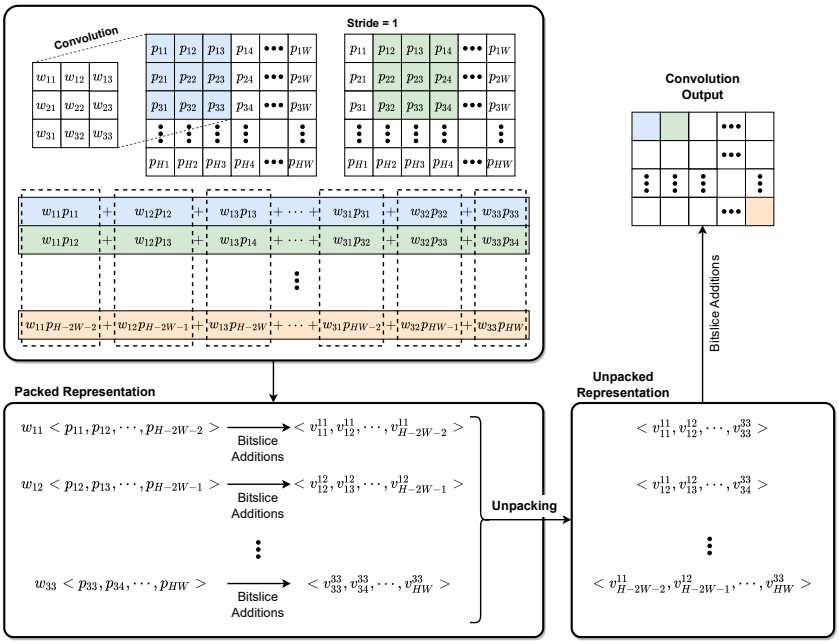


Fig. 5: Data Packing with Bitslicing for FHE-Conv2D Workflow

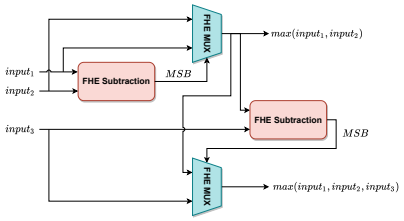


Fig. 6: FHE-MaxPool2D Logic

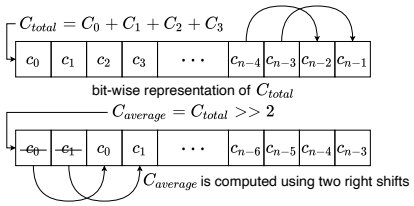


Fig. 7: FHE-AvgPool2d (Logical Bitwise Circuitry)

performed by bit shifts on the ciphertext vectors. As (2×2) AvgPools are very popular, computing AvgPool2D is just division by 4 (i.e. 2 Right Shift(RS) operations. Thus we get our simplified FHE circuit mappings for AvgPool2D calls. As demonstrated in Fig. 7, (C_1, C_2, C_3, C_4) are four ciphertexts which will be fed into a 2×2 AvgPool2D module. Our circuit first adds these four ciphertexts and then obtains the $AvgPool2D$ result by performing 2 logical right bit shift $(2RS)$ operations. Likewise, MaxPool2D FHE-circuit mappings could be realized by an encrypted multiplexer (FHE-MUX). The circuit selects the number based on the MSB difference of two input numbers. Our circuit, in this regard, is a cascade of FHE MUXs. For example, if we apply MaxPool2D from 3×3 kernel, then our circuit is a cascade of (3×3)-1=8 MUXs. Each MUX selects the maximum of two numbers based on sign of difference of the numbers. An illustration for the MaxPool is shown in Fig. 6. At this point, we note that the AvgPool2D is significantly lightweight compared to the MaxPool. Therefore, in both of our models, we utilize the AvgPool2D circuit.

Encrypted Activation Functions: An Activation Function decides whether a neuron should be activated or not, to be specific, whether the neuron's input to the network is important or not in the process of prediction using simpler mathematical operations. Since, linear activations are not much effective in DNN or CNN scenario, we have mostly focused on efficient FHE counterpart design for nonlinear activations.

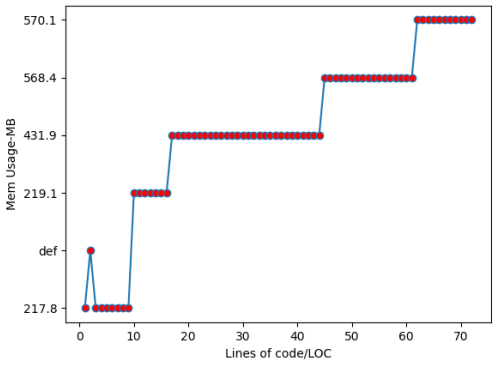
Among nonlinear activations, ReLU (Rectified Linear Unit) is one of the most used functions in deep learning applications. As discussed in section II, existing encrypted ML architectures are either based on LHE or some approximations of FHE schemes with some inherent limitations that does not allow to incorporate ReLU in encrypted domain. That is the reason why the existing schemes mostly use either square activations or sign activations. To elaborate the reason of this limitation further, we revisit the basic operations required to implement ReLU activation.

ReLU $(f(x))$ as an activation that selects the positive part of its argument as $f(x) = max(0, x)$, where x is the input to a neuron. Realization of maximum function in homomorphic domain requires encrypted decision making, that is not feasible with LHE computations [40], [41]. In our FHE based implementation, we realize the encrypted decision making with encrypted multiplexer considering the FHE subtraction result as selection input to choose between encrypted 0 and encrypted x.

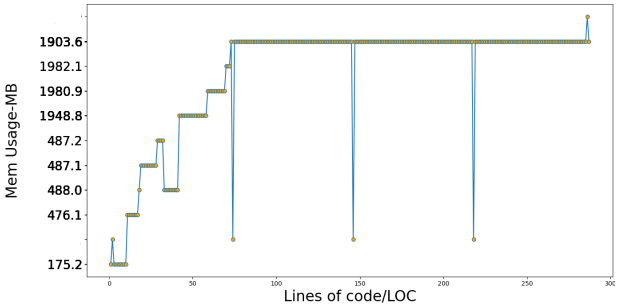
However, other non-linear activations like *Softmax* or *tanh* include the standard exponential function computation on en-

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:52:22 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2022.3204909



(a) Conv2D Design-Memory Profiles (with Bitslicing)



(b) Memory Profiles of only the *Dense Layer* of Conv2D+AvgPool+Dense Inference Pipeline (with Bitslicing)

Fig. 8: Memory Profiling Statistics of Bitsliced Inference (Our Method)

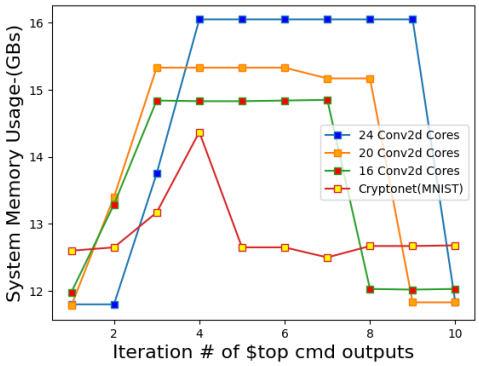


Fig. 9: Memory Profiles of SHE and Cryptonet Baselines

TABLE VII: Speedup for bitslicing DNN architecture wrt. non-bitslicing.

Experiment Setup	NN-Inference Time (Hours)
PyFHE	120
NuFHE(without Bitslicing)	80
NuFHE(Bitsliced)	2.15

crypted input vectors. As explained in [42], exponent computation becomes infeasible with the encrypted power term considering our underlying unencrypted processors. However, this limitation can be outdone in ML processing with prior domain knowledge, where the maximum value of such input vectors are known [42].

V. Experiments and Results

In this section, we present the timing and memory overhead of our proposed encrypted inference architecture. We evaluate

TABLE VIII: Bitsliced CNN timing profiles (MNIST)

(28×28) MNIST Data	Time in Minutes (=Hours)
Conv2D	$139.78(= 2.32)$
AvgPool2D	$11.53(= 0.2)$
Conv2D+AvgPool2D	$151.31(= 2.52)$
Conv2D+AvgPool2D+Dense	$268.21(= 4.47)$

TABLE IX: Bitsliced CNN timing profiles (Mendeley)

$(224 \times 224 \times 3)$ Mendeley Data	Time (Estimated Hours)
Conv2D	108
AvgPool2D	30
Conv2D+AvgPool2D	138
Conv2D+AvgPool2D+Dense	140

all our experimental studies on an Ubuntu 18.04LTS $64-bit$ platform with $256GB$ RAM and $24GB$ GPU memory support. All underlying FHE blocks are supported by existing FHE encryption library NuFHE [16] and PyFHE [43], which is another version of TFHE for performance comparison. NuFHE [16], which is a GPU powered Torus FHE encryption library implemented in Python, where PyFHE does not have any GPU support.

We start with a simple DNN architecture which includes three layers: two layers of 50 neurons and a final fully connected layer of 10 neurons. Table VII compares the bitsliced implementation of a simple DNN architecture with non-bitsliced versions of the same architecture. In bitsliced implementation we first break down our target architecture to proposed FHE counterparts of Conv2D, AvgPool2D, Dense. These basic blocks are accelerated by underlying bitsliced adder functions developed on top of NuFHE homomorphic blocks. As shown in Tables VII, compared to non-bitsliced inference with timing overhead of 80 hours, our bitslicing based software wrappers can accelerate the same task to 2.15 hours in case of DNN implementation.

Table VIII and IX shows the timing profiles of the bitsliced implementation of the CNN architecture for both MNIST Digit Classification [36] and Mendeley use-case [37]. The reference architectures for both the use-cases have been elaborated in Table V and in Figure 3. We provide an estimated timing requirement for execution of this same CNN architecture on the Mendeley Bridge dataset [37] as shown in the Table IX. The dimensions of this dataset is $224 \times 224 \times 3$. We note that this dataset is more complex than other representative datasets, such as CIFAR10 [44] (dimension $32 \times 32 \times 3$), in terms of input dimension which is a critical factor in FHE

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:52:22 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2022.3204909

TABLE X: SHE/Bitslicing Memory Profile Baseline Comparisons

Number of Conv2D Operations	Memory Overhead (SHE with TFHE)	Memory Overhead(NuFHE-Bitslicing)
30	6GB	$<$ 200MB
50	9GB	$<$200MB
100	18GB	$<$400MB
676 (MNIST)	$100GB^*$	570MB

as it drastically increases the computation. As per Table VIII the bitsliced Conv2D takes about 2.3 hours while complete evaluation of an MNIST image takes about 4.5 hours and it may take around 140 hours for Mendeley (Table IX).

Memory Profiling of Bitsliced Inference

Recent concerns about security and privacy guarantees in IoT systems have drawn interests about secure computation in the IoT domain. However, huge memory consumption and computation overhead are two prime bottlenecks to implement HE schemes on resource constrained embedded platforms and this is an important recent research direction [45]. With this motivation, we have performed layer-by-layer memory profiling of both our DNN and CNN architectures as shown in the Figures 8b and 8a. As depicted in Figure 8a, bitslicing supported CNN framework is quite memory efficient (about $700MB$). The DNN framework, which takes about $1.9GB$ also fits into the modest memory bounds of Edge devices i.e. 2-4GB as highlighted in figure 8b.

Furthermore, we have compared our memory profiles to the state-of-the-art encrypted inference approaches. We benchmark our bitsliced design pipeline with an open source implementation [2] due to [7]. This approach provides an implementation of scalar products with small examples e.g. 8 ciphertexts. We use this codebase to estimate their memory growth and we highlight the same in Table X. To do this, we approximate their implementation i.e. 8 scalar products as equivalent to a single Conv2D operation (the reason being with a 3×3 kernel, Conv2D is just 9 scalar products which the concerned framework replace with logical bit-shifts). We check experimentally that a single Conv2D operator amounts to around $180MB$ memory overhead. Then, we run multi-threaded implementation of this Conv2D block and observe the memory explodes proportionally as highlighted for several thread values $(16, 20, 24)$ in Figure 9. Therefore, as Table X depicts, an evaluation of 50 convolutions amounts to about 9GBs. We also evaluate memory footprints of Cryptonets [4], another pioneering approach by Microsoft for secure CNN inference. This approach exhausts 2-3 GB memory for running a simple MNIST pipeline with a Conv2D stacked with couple of Dense layers. Research report in [5] also points out that most encrypted CNN approaches amounts to atleast few GB of RAM usage. Table XI shows the merits and demerits of our proposed architecture with some state of the art encrypted inference models [4], [5], [7], [14]. Comparisons are done on the basis of power of homomorphic computations (leveled or fully homomorphic), activations supported, associated memory overheads and possibility of decryption errors. It is to note that existing inference models may outperform in terms of timing requirement, but our proposed decryption error free model has its own significance in terms of better accuracy without any approximation and hugely reduced memory overhead. After memory profiling all these baselines, we would like to conclude that the proposed scheme hugely optimizes the memory footprint and this aspect perhaps cements our model's potential for low resource edge platforms.

[2]https://github.com/qianlou/SHE

VI. Conclusion and Future Research

In this work, we realize encrypted CNN suitable for real-world DL-based applications. In this paper, we provide several key insights. Performance and blow up in ciphertext size are two major bottlenecks, which we address with architectural tricks, such as quantization and bitslicing. A potential future work in this regard is to further optimize the inference timing for a larger spectrum of real-world applications. Another direction of future research is to adopt the proposed optimizations beyond Boolean FHE.

Acknowledgements: The work was partially supported by the projects sponsored by SERB SUPRA DST Govt of India and US-India cooperation in higher education with Univ of Nebraska at Omaha.

References

[1] P. Nicholson, "Aws hit by largest reported ddos attack of 2.3 tbps," 2020, https://www.a10networks.com/blog/aws-hit-by-largest-reported-ddos-attack-of-2-3-tbps/.

[2] D. Ramel, "Epic ddos fail: Azure cloud fends off 'largest attack ever reported in history'."

[3] E. Root, "Cryptominers threaten gcp virtual servers," 2022, https://www.kaspersky.com/blog/attacks-on-google-cloud-platform/43312/.

[4] N. Dowlin, R. Gilad-Bachrach, K. Laine, K. Lauter, M. Naehrig, and J. Wernsing, "Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy," in *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ser. ICML'16. JMLR.org, 2016, p. 201–210.

[5] A. Brutzkus, O. Elisha, and R. Gilad-Bachrach, "Low latency privacy preserving inference," in *International Conference on Machine Learning*, 2019.

[6] N. Jain, K. Nandakumar, N. K. Ratha, S. Pankanti, and U. Kumar, "Efficient CNN building blocks for encrypted data," *CoRR*, vol. abs/2102.00319, 2021.

[7] Q. Lou and L. Jiang, "She: A fast and accurate deep neural network for encrypted data," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc.

[8] C. Gentry, "Computing on encrypted data," in *International Conference on Cryptology and Network Security*. Springer, 2009, pp. 477–477.

[9] L. Ducas and D. Micciancio, "Fhew: Bootstrapping homomorphic encryption in less than a second," in *Advances in Cryptology – EUROCRYPT 2015*, E. Oswald and M. Fischlin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 617–640.

[10] I. Chillotti, N. Gama, M. Georgieva, and M. Izabachène, "TFHE: Fast fully homomorphic encryption over the torus," *J. Cryptol.*, vol. 33, no. 1, p. 34–91, jan 2020.

[11] J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," Cryptology ePrint Archive, Report 2016/421, 2016, https://ia.cr/2016/421.

[12] B. Li and D. Micciancio, "On the security of homomorphic encryption on approximate numbers," in *Advances in Cryptology - EUROCRYPT 2021 - 40th Annual International Conference on the Theory and Applications of Cryptographic Techniques*, ser. Lecture Notes in Computer Science, A. Canteaut and F. Standaert, Eds., vol. 12696. Springer, 2021, pp. 648–677.

[13] U. Crypto, "Ckks key recovery," https://github.com/ucsd-crypto/CKKSKeyRecovery, Mar 2021.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:52:22 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. This is the author's version which has not been fully content may change prior to final publication. Citation information: DOI 10.1109/TCAD.2022.3204909

TABLE XI: Comparison with Existing Approaches

Framework	Library	Activations (Exact/Approximate)	Activations Supported	Memory Overhead	Erroneous Decryption
Cryptonets	LHE	Exact	Square	$>$ 100GB (CIFAR-10)	No
Lola-Cryptonets	LHE	Exact	Square	$2-3$GB (MNIST)	No
FHE-DiNN	FHE	Approximated	Signed	–	Yes
SHE (estimated for deeper nets)	LHE	Approximated	ReLU	100 GB (MNIST estimated)	No
Our Approach	FHE	Exact	All	$< 700MB$ (MNIST)	No

[14] F. Bourse, M. Minelli, M. Minihold, and P. Paillier, "Fast homomorphic evaluation of deep discretized neural networks," in *Annual International Cryptology Conference*. Springer, 2018, pp. 483–512.

[15] ZAMA, "zama.ai/concrete," https://github.com/zama-ai/concrete, 2021.

[16] NuFHE, "Nufhe, a gpu-powered torus fhe implementation." [Online]. Available: https://nufhe.readthedocs.io/en/latest/

[17] N. Gama, "Tfhe." [Online]. Available: https://github.com/tfhe/tfhe

[18] E. Biham, "A fast new DES implementation in software," in *Fast Software Encryption, 4th International Workshop, FSE '97*, E. Biham, Ed., vol. 1267. Springer, 1997, pp. 260–272.

[19] J. Lee, H. Kang, Y. Lee, W. Choi, J. Eom, M. Deryabin, E. Lee, J. Lee, D. Yoo, Y. Kim, and J. No, "Privacy-preserving machine learning with fully homomorphic encryption for deep neural network," *IEEE Access*, vol. 10, pp. 30 039–30 054, 2022.

[20] R. L. Rivest, L. Adleman, M. L. Dertouzos *et al.*, "On data banks and privacy homomorphisms," *Foundations of secure computation*, vol. 4, no. 11, pp. 169–180, 1978.

[21] T. E. Gamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," in *Advances in Cryptology, Proceedings of CRYPTO '84, Santa Barbara, California, USA, August 19-22, 1984, Proceedings*, ser. Lecture Notes in Computer Science, vol. 196. Springer, 1984, pp. 10–18.

[22] Z. Brakerski, C. Gentry, and V. Vaikuntanathan, "(leveled) fully homomorphic encryption without bootstrapping," in *Innovations in Theoretical Computer Science 2012, Cambridge, MA, USA, January 8-10, 2012*, S. Goldwasser, Ed. ACM, 2012, pp. 309–325.

[23] A. Al Badawi, C. Jin, J. Lin, C. F. Mun, S. J. Jie, B. H. M. Tan, X. Nan, K. M. M. Aung, and V. R. Chandrasekhar, "Towards the alexnet moment for homomorphic encryption: Hcnn, the first homomorphic cnn on encrypted data with gpus," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 3, pp. 1330–1343, 2021.

[24] E. Hesamifard, H. Takabi, and M. Ghasemi, "Cryptodl: Deep neural networks over encrypted data," *ArXiv*, vol. abs/1711.05189, 2017.

[25] B. Moons, D. Bankman, and M. Verhelst, *Embedded Deep Learning: Algorithms, Architectures and Circuits for Always-on Neural Network Processing*, 01 2019.

[26] I. Chillotti, M. Joye, and P. Paillier, "Programmable bootstrapping enables efficient homomorphic inference of deep neural networks," Cryptology ePrint Archive, Report 2021/091, 2021, https://ia.cr/2021/091.

[27] J.-W. Lee, H. Kang, Y. Lee, W. Choi, J. Eom, M. Deryabin, E. Lee, J. Lee, D. Yoo, Y.-S. Kim *et al.*, "Privacy-preserving machine learning with fully homomorphic encryption for deep neural network," *IEEE Access*, vol. 10, pp. 30 039–30 054, 2022.

[28] C. Gentry, "A fully homomorphic encryption scheme." [Online]. Available: https://crypto.stanford.edu/craig/craig-thesis.pdf

[29] O. Regev, "On lattices, learning with errors, random linear codes, and cryptography," vol. 56, no. 6, 2009.

[30] C. Gentry, A. Sahai, and B. Waters, "Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based," *IACR Cryptol. ePrint Arch.*, vol. 2013, p. 340, 2013.

[31] Concrete, "Zama." [Online]. Available: https://www.zama.ai/concrete-framework

[32] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[33] "Data encryption standard (des)." NIST, 1995.

[34] "Tensorflow model optimization." [Online]. Available: https://www.tensorflow.org/model_optimization/guide/quantization/training

[35] NVIDIA, "nvidia/fastertransformer," https://github.com/NVIDIA/FasterTransformer, 2021.

[36] Y. LeCun, C. Cortes, and C. Burges, "Mnist handwritten digit database," *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist*, vol. 2, 2010.

[37] C. F. Ozgenel, "Concrete crack images for classification," https://data.mendeley.com/datasets/5y9wdsg2zt/2, 2019.

[38] "Quantized tflite model gives better accuracy then tf model." [Online]. Available: https://stackoverflow.com/questions/64030764/quantized-tflite-model-gives-better-accuracy-than-tf-model

[39] V. Praveen, "Improving int8 accuracy using quantization aware training and the nvidia tao toolkit," https://developer.nvidia.com/blog/improving-int8-accuracy-using-quantization-aware-training-and-tao-toolkit/#:~:text=Conclusion,compared\%20to\%20FP16\%20or\%20FP32.

[40] A. Chatterjee and I. Sengupta, "Translating algorithms to handle fully homomorphic encrypted data on the cloud," *IEEE Transactions on Cloud Computing*, vol. 6, no. 1, pp. 287–300, 2018.

[41] A. Chatterjee, M. Kaushal, and I. Sengupta, "Accelerating sorting of fully homomorphic encrypted data," in *Progress in Cryptology - INDOCRYPT 2013 - 14th International Conference on Cryptology in India, Mumbai, India, December 7-10, 2013. Proceedings*, ser. Lecture Notes in Computer Science, G. Paul and S. Vaudenay, Eds., vol. 8250. Springer, 2013, pp. 262–273.

[42] A. Ghosh, A. Raj, and A. Chatterjee, "Encrypted operator design with domain aware practical performance improvement," in *Proceedings of the Seventh International Conference on Mathematics and Computing - ICMC 2021*, D. Giri, K. R. Choo, S. Ponnusamy, W. Meng, S. Akleylek, and S. P. Maity, Eds., vol. 1412. Springer, 2021, pp. 93–107.

[43] "pyFHE github repository," Jan 2020. [Online]. Available: https://github.com/virtualsecureplatform/pyFHE

[44] A. Krizhevsky, "The cifar10 dataset," http://www.cs.toronto.edu/~kriz/cifar.html, 2009.

[45] D. Natarajan and W. Dai, "Seal-embedded: A homomorphic encryption library for the internet of things," *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, vol. 2021, no. 3, pp. 756–779, 2021.

Soumik Sinha is currently pursuing his PhD degree at Indian Institute of Technology Kharagpur. His research interest lies at the confluence of System Security and AI.

Sayandeep Saha is currently a Post Doctoral Researcher at NTU Singapore. His research interests include formal methods and hardware security.

Manaar Alam is currently a Post Doctoral Researcher at NYU Abu Dhabi. His research interests lie at the confluence of Deep Learning and Security.

Varun Agarwal received his Bachelors in Technology from Delhi Technological University, Delhi, India in 2022. His research interest includes image cryptography, hardware security and homomorphic encryption.

Ayantika Chatterjee is an Assistant Professor at Indian Institute of Technology Kharagpur. Her research interest includes Security, Embedded Systems and Encrypted Computation.

Anoop Mishra is a PhD Candidate at University of Nebraska Omaha, working in different areas of Artificial Intelligence, Internet-of-Nano-things, Edge Computing etc.

Deepak Khazanchi is a Professor of Information Systems and Quantitative Analysis at the University of Nebraska Omaha.

Debdeep Mukhopadhyay is a Professor in the Department of Computer Science and Engineering, IIT Kharagpur. His research interests include cryptography, VLSI of cryptographic algorithms, hardware security, and side channel analysis.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 25,2022 at 04:52:22 UTC from IEEE Xplore. Restrictions apply.

2022 2nd Asian Conference on Innovation in Technology (ASIANCON)
Pune, India. Aug 26-28, 2022

Feasibility Study and Power Flow Management of DC Bus for Household Loads

Anupam
Electrical Engineering Department
Delhi Technological University
Delhi, India
anupamsharma478@gmail.com

Mayank Kumar, *SM IEEE*
Electrical Engineering Department
Delhi Technological University
Delhi, India
mayankkumar@dtu.ac.in

Suman Bhowmick
Electrical Engineering Department
Delhi Technological University
Delhi, India
suman.bhowmick@dce.ac.in

2022 2nd Asian Conference on Innovation in Technology (ASIANCON) | 978-1-6654-6851-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ASIANCON55314.2022.9908686

***Abstract*—This paper proposes the technique to standardize 48V DC power line for homes and offices, energy storage systems (ESSs) are also integrated with solar photovoltaic (SPV) sources to control the effect of intermittency and periodicity. Battery acts as an energy buffer when SPV system could not meet the requirement of load demand. It is well known that SPV provides DC power and stores in ESS, so both of them are integrated to the DC bus. With the help of power electronics step down converter, the 48V DC bus is stepped-down to 24V which is used to power the low voltage devices like laptop, desktop, LEDs and TVs. As some appliances works on AC input, thus, with the help of power inverter 48V DC boosted up to 230V AC to meet the required voltage level. This paper further explains the model in Simulink which consists SPV array system with battery energy storage device. The DC bus uses DC-DC converter, an inverter and a bidirectional converter for charging/discharging of ESS. A Maltab/Simulink platform is used to validate the feasibility of 48V DC power line with ESS under varying irradiances.**

Keywords— Battery energy storage system (BESS), bidirectional converter (BDC), MPPT, power flow management, solar pv array (SPV).

I. INTRODUCTION

The idea of DC distribution has taken the shape because of large power demand due to technological advancement in the world which could not be meet by the conventional energy sources, so the development in regards of power electronics is redirected, which eases the application of renewable energy sources [1]. Solar are considered to be the more efficient for energy conservation among all of the renewable energy sources due to their free availability, pollution free and located easily as per demand without any issues related to size and configuration. SPV directly generate DC power which can easily be stored in the battery energy storage system [2], [3].

In today's arena, power used by most of the appliances are DC in nature. So renewable energy first converted into AC and making it expedient for supplying power into homes/offices buildings as per present distribution system, but now these days many of the modern appliances manufactured internally on DC power so the system requires two step conversion (DC→AC→DC) which makes the losses of energy up to 30% [4], [5]. To eliminate these energy losses, we need to transform present distribution system from AC to DC (i.e. linking of DC loads directly to DC bus), this makes the system more economical and efficient [6].

SPV are modelled here and corresponding parameters are described in TABLE I. MPPT perturb and observe control algorithm is used here for extraction of maximum energy from varying solar radiation [7]. Intermittent and periodic nature of solar photovoltaic source is one of the major drawback, so battery energy storage support is the requirement for regulating output power and avoid the fluctuations.

For efficient DC distribution system, voltage level plays a major role for switching surge at high voltage level and cable losses at low voltage level. Transition of existing AC system to DC system can be done with minimum modification in appliances. The adaptation of conventional AC appliances in DC system depends upon the chosen DC voltage level.

The detail work in this research is elaborated as follows. Section II describes the system model description of solar energy modelling with MPPT algorithms. Section III describes charging algorithm and load design. Section IV describes energy management method for emphasizing all the objectives. Section V represents simulation results follows by conclusions in section VI.

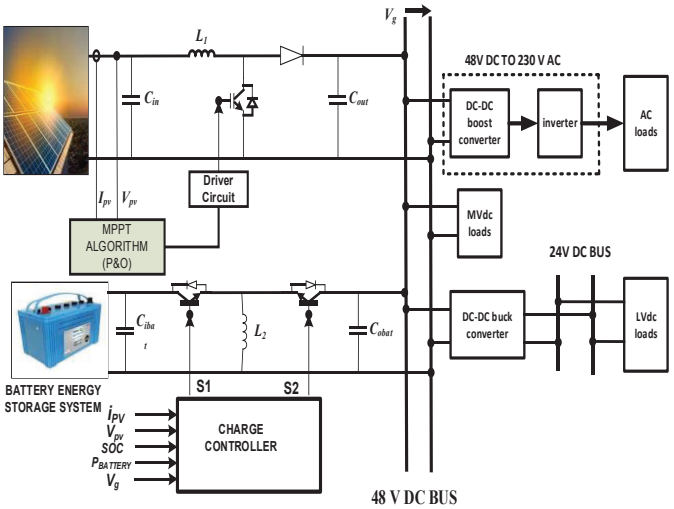


Fig.1.SPV integrated system with BESS and household loads

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:09:41 UTC from IEEE Xplore. Restrictions apply.

II. System Model Description

A. Modelling of Photovoltaic cell

Fig.2 represents the modelling of photovoltaic cell. PV cell is connected in series and parallel combination and considered under standard test condition (i.e. 1000W/m^2 and ambient temperature 50°C). In this simulation, soltech 1STH-215-P solar panel is used. The following equations has developed from fig.2 which describes the I-V characteristics of the PV solar cell.

$$I_{ph} = I_d + I_{R_{sh}} + I_{pv} \quad (1)$$

$$I_{pv} = I_{ph} - I_d - I_{R_{sh}} \quad (2)$$

$$I_{pv} = I_{ph} - I_o e[\frac{q(V_{pv} + I_{pv}R_s)}{nKT}] - \frac{V_{pv} + I_{pv}R_s}{R_{sh}} \quad (3)$$

Where I_{pv} is PV current of solar cell (A); I_{ph} is the light generated current (A); I_o is the diode reverse saturation current (A); R_s is the series resistance of cell (ohms); R_{sh} is shunt resistance of cell (ohms); V_{pv} is the voltage of solar cell; K is the Boltzmann's constant; T is the temperature in kelvin; q is the electron charge [8].

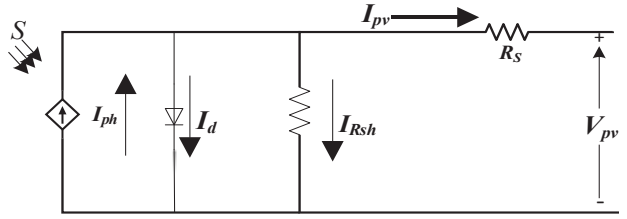


Fig.2.Circuit model of PV cell

TABLE I. Electrical Characteristics of PV Panel (Soltech 1STH-215-P) at STC (1000 W/M^2 and 50°C)

Parameters	Value
N_P (parallel strings of PV)	2
N_S (series strings of PV)	1
V_{OC} (voltage at open circuit)	36.3V
I_{SC} (current at short circuit)	7.84A
V_{MP} (MPP Voltage)	30V
I_{MP} (MPP Current)	7.35A
P_{MP} (MPP Power)	213.15W

B. Maximum power point tracking algorithm

The MPPT is defined as an electronic device which draws the maximum power from the solar module. The proposed work has used the P&O technique for the MPPT. In the P&O technique, small perturbation is seen which causes the variation of power in PV module. The SPV power output is measured in periodic manner and compared with previous powers. The main concept of this algorithm is to analyze and compare the output power and voltages of the PV array with its corresponding previous values. With the help of PV module the voltage will be either decreased or increased in order to check whether the power has decreased or increased. The power at time 'K' is compared with time 'K-1' which generates the signal, from this signals the power is governed for the net instant towards maximum power point. This procedure observe the output voltage and power of the PV array that perturbs the signal. There will be increment and decrement of the voltage signals by the constant amount which further sent to the boost converter. The direction of signals further decides on the observation and comparison with the SPV power output. The obtained value is further compared with the reference value at the MPPT point, which is based on the direction of the signal. The signal will be ΔV when the output load power increases and it will -ΔV when output load power decreases [10].

C. Lithium-ion Battery

A lithium-ion battery has opted for the power compensation.

Fig.3 represents the model of Lithium-ion battery with battery charge controller. Batteries are connected in series-parallel in order to increase the voltage and current for large storage application. The voltage (V_o) of battery can be represented as the sum of voltage across internal resistors and internal voltage represented by equation (4).

$$V_0 = V_{B_{\min}} + (V_{B_{\max}} - V_{B_{\min}})\sqrt{\frac{SOC}{100}} \quad (4)$$

For the safety operation, the battery has to be kept within the minimum and maximum voltage (V_{Bmin} and V_{Bmax}). The battery charged by constant voltage V_{Bmax} till the voltage reaches to its maximum value. Meanwhile, the charging current starts decreasing. The SOC and internal resistances determine the battery current. The charging/discharging process of lithium-ion battery represented by equation (5),(6).

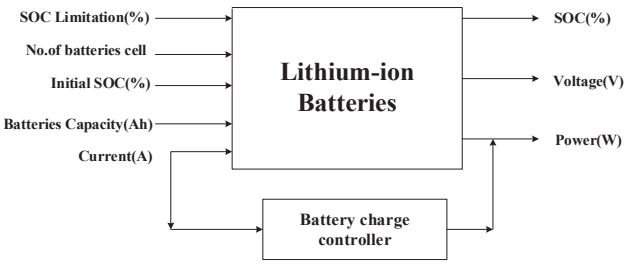


Fig.3 Structure of Li-ion battery system

$$C_{ch_batt}(t) = (t-1)(1-\sigma)C_{_batt} - \left[\frac{P_b(t)}{\eta_{inv}} - P_{pv}(t)\right] \quad (5)$$

$$C_{dis_batt}(t) = (t-1)(1-\sigma)C_{_batt} - \eta_{batt}\left[P_{pv}(t) - \frac{P_b(t)}{\eta_{inv}}\right] \quad (6)$$

Where, C_batt(t) and C_batt(t-1) are the available capacities (Wh) at t and t-1 respectively, σ is the self-discharge rate, P_{pv}(t) is generated energy by SPV system. P_L(t) is the load power demand at time t, η_{inv} is efficiency of inverter and η_{batt} is charging/discharging efficiency of the battery.

D. Bidirectional converter and its control topology

Fig.4 shows the non-isolated DC-DC converter which is applied for battery controller. The bidirectional converter sets for charging/discharging based upon the load power demand. In case of charging the battery, the switch S_2 is on and power flows from DC bus to battery. But when the load power demand is high during night time and SPV power is negligible, then battery starts discharging. In this case switch S_1 will be on that will discharges the battery and power starts flowing from the DC bus to battery [11].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:09:41 UTC from IEEE Xplore. Restrictions apply.

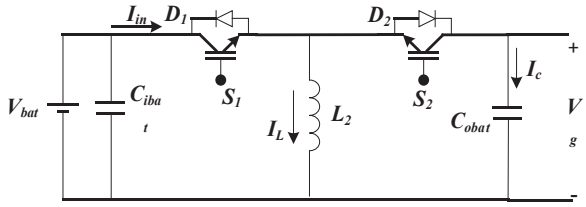


Fig.4 Circuit diagram of bidirectional converter

Fig.5 shows the control techniques of bidirectional converter. From fig.5 it is observed that battery current reference signal (I_{batt}) compared with actual battery current signal (I_{batt_act}). The error signal passes through PI controller. The PI controller output is compared with voltage signal (V_{ref}) to generate pulses for battery. In the bidirectional converter AND logic gates are used for switching operation. When SPV produces surplus power, its supplies power to AC/DC load. In this case BDC is in on mode (mod=1), that operates in buck converter mode for charging the battery. When SPV is unable to fulfil the load demand, then fully charged battery device will suffice the load power requirement. In this case BDC will operates in boost mode (mod=0). Hence BDC will discharges the power to cater load side requirement.

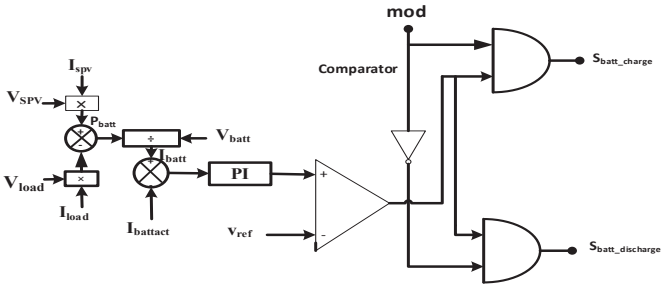


Fig.5 Control strategy of bidirectional DC-DC converter

III. CHARGING ALGORITHM AND LOAD DESIGN

A. Battery Charging Algorithm

Lithium ion batteries follows constant current - constant voltage (CC/CV) charging algorithm. Under the CC/CV charging algorithm, there exist three step charging process. In the charging algorithm of (CC/CV), battery is charge by constant current till the voltage of battery storage rises to maximum preset voltage ($V_{\max\, pre_set}$), after then charging voltage will held constant at V_{pre_set} . Further, charging current reduces exponentially after reaching at maximum preset voltage. Three steps of charging process is considered for battery charging algorithm, at very first step battery is at initial condition, and here the temperature and open circuit voltage (OCV) are checked. If both of them are in normal range then OCV is lower than preset cutoff voltage (V_{cutoff}). Here the battery is charged by the trickle charging (TC) method. In second step battery voltage exceeds V_{cutoff} , here the CC mode starts charging the battery. In the second mode, the charging current is chosen such that the lithium ion batteries range varies from 0.5C to 3.2C. In the third step battery voltage charges to preset value of V_{pre_set} . After third step, battery charging process switches to CV mode and corresponding charging current starts reducing correspondingly. The termination of charging current is done either by minimum charging current ($I_{\min} < 0.1C$) or the maximum charging time ($t_{\max} < t_{pre-set}$).

B. Loads appliances and its design

In this section we have taken two types of loads, DC loads which consists of low voltage residential appliances that runs with the help of stepping down the voltage of DC bus. The other one is considered as AC loads for high voltage residential appliances (i.e. Refrigerator, air-conditioner etc.). From the conventional consideration the compressor of refrigerator and air-conditioner composed with an induction motor which is ac in nature that runs with fixed speed. The rotational speed is working at particular input frequency and compressor works with fixed speed which produces high impact on compressor mechanics and its control mechanism.

Meanwhile, the inverter based technology for air-conditioner and refrigerator has invented now, which depends upon power electronics components for handling its driving equipment's. The higher loads appliances are made to work on 230 V AC, lower and medium voltage residential appliances are made to work on 24 V and 48V DC respectively. Table II represents the respective loads of household.

TABLE II. LOAD CONSIDERED FOR HOUSEHOLD APPLIANCES

SR. NO.	Appliances	Quantity	Power Required(W)	No.of hours
1.	LED BULB(20W)	5	100	5
2.	Laptop (50W)	1	50	2
3.	Fan (40W)	3	120	8
4.	TV (100W)	1	100	4
5.	Refrigerator (100W)	1	100	8
	TOTAL		**470W**	

Overall energy consumption from the load appliances:

$$Wh_{load} = (20\times5\times5)+(50\times1\times2) + (40\times3\times8)+(100\times1\times4)+(100\times1\times8)$$
$$=(500+100+960+400+800)=2760\text{Wh}$$

The nominal dc bus voltage is used in the proposed manuscript is taken as $V_{bus} = V_g = 48V$

Autonomy days (N) =1 day

Efficiency of battery $\eta = 85\%$

Depth of discharge of battery (DOD) = 70%

$$Battery\ capacity = \frac{Wh_{load}\times(N+1)}{\eta\times DOD\times V_g} = \frac{2760\times2}{0.85\times0.7\times48} \approx 193Ah$$

So, for safety criteria battery capacity to be taken as 200Ah.

$$\text{Peak load demand} = (100+50+120+100+100) = 470W$$

$$\text{Maximum load current} = \frac{470}{48} = 9.79A$$

Hence from the above calculations the discharging current is taken to be 9.79A for transferring surplus PV

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:09:41 UTC from IEEE Xplore. Restrictions apply.

power to battery without disturbing the bus voltage during the high insolation period and low load power demand,

$$\text{Now, C-rating of battery}\left(\frac{200Ah}{9.79}\right) \approx C/20$$

Hence, battery will provide 10A of current continuously at full load for 20 hours in the absence of PV array with 200Ah battery capacity.

IV. ENERGY MANAGEMENT METHOD

Fig.6 represents the step by step power balancing for various operational scenario, if SPV power greater than the load power demand and SOC of battery is lower than 80%, then battery will be operating in charging mode. Moreover, if SOC lies between 20% and 80% and SPV power lower than the load power demand, then battery discharges to fulfill the load power necessity. The MPPT will be turned off, if SOC >80%. This below cases represents the energy management of SPV+BESS based standalone DC grid system [12].

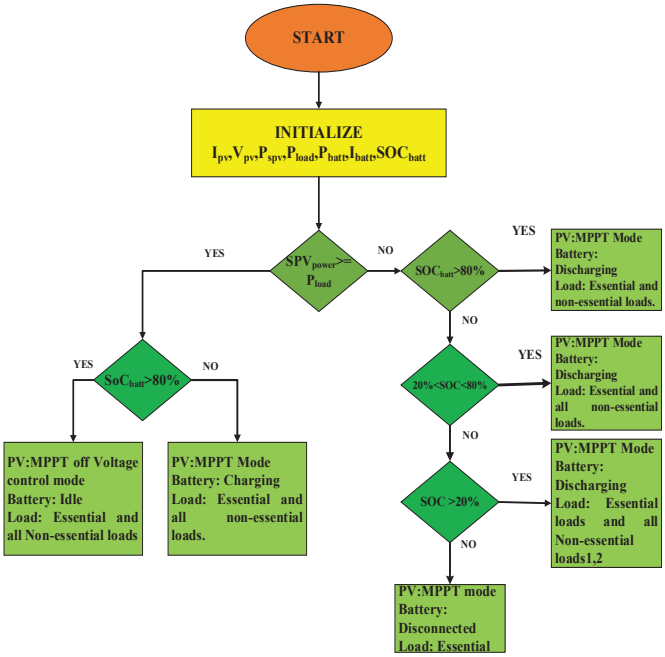


Fig.6 Energy management scheme of DC grid

Case I: $SPV_{power} > P_{load}$ **and** $SOC_{batt}^{min} < SOC_{batt} < SOC_{batt}^{max}$

If the solar PV array power exceeds the load power demand and if SOC of battery is less than its maximum threshold value i.e. SOC_{batt}^{max} . The specified condition checked by the control algorithm and the control algorithm provides control signal to charge the battery. In this case all the essential and non-essential loads are powered by solar PV and battery. The battery will be charged to its threshold value.

Case II: $SPV_{power} > P_{load}$ **and** $SOC_{batt} = SOC_{batt}^{max}$

In this case, if SPV_{power} is more than the load power demand and battery SOC_{batt} is at maximum threshold value i.e. SOC_{batt}^{max} . The battery gets disconnected from the system and only solar PV array alone will meet the requirement of entire load demand.

Case III: $SPV_{power} = P_{load}$ **and** $SOC_{batt} = SOC_{batt}^{max}$

In this case, the generated power of SPV equals to the load power demand, hence the SOC of BESS equals to the maximum threshold value of battery. In this case BESS not considered for the purpose of optimal operation. The SPV_{power} alone will satisfy the entire load demand.

Case IV: $SPV_{power} < P_{load}$ **and** $SOC_{batt} > SOC_{batt}^{min}$

In this case, if the SPV array power is less than the load power demand and if battery SOC is greater than its minimum threshold value i.e. SOC_{batt}^{min} . Then (P_{spv} + BESS) both are combined and satisfy the power of load demand. So, both (P_{spv} + BESS) delivers power to essential and non-essential loads on the priority basis and Hence the battery SOC gradually reduces to its lower value i.e. SOC_{batt}^{min} .

Case V: $SPV_{power} < P_{load}$ **and** $SOC_{batt} < SOC_{batt}^{min}$

In this case if the power generated from SPV array is less than load power demand, and if the SOC of battery is less than the minimum threshold value i.e. SOC_{batt}^{min} . Then both (P_{spv} + BESS) delivers power to essential and non-essential loads. The non-essential loads will be disconnected according to priority basis.

In all the above five cases the battery SOC limits are taken between 20% and 80%. Furthermore, it is varied with connected loads.

A. Control objectives and scope

In the DC grid structure, three levels of control have been proposed: primary level, secondary level, and tertiary level as shown in fig.8. Voltage loop control, current loop control, MPPT control and SOC estimation comes under primary level control. Its objective is to further regulate local voltages and currents. It also controls the breakers, switches (on/off protection functions) and load control. Whenever high rating loads are connected or disconnected from the DC grid, voltage changes are caused by the primary control and hence the system will be protected from under or overvoltage. To resolve this concern, a secondary control strategy is considered for voltage quality enhancement and voltage unbalance compensation. It's also helps in the improvement of performance. This also covers voltage control for DC bus voltage regulation. The economic dispatch, optimal power flow and optimal energy scheduling are all controlled by a tertiary control scheme. Tertiary control is regarded as EMS, which helps in accurate power control among source and load. By using this control multi objective function is specified using innovation such as power generation, demand forecasting, weather forecast, energy storage and energy grid prices. The system will become more adaptive and efficient as a result of the above control and objective functions. DC Grid EMS is considerable as multi-objective system which deals with technical, environmental and economical issues. Moreover, its main objective is to optimize the energy scheduling operation and reliability of the system in both standalone as well as in battery connected mode. For the controlling purposes, it is mandatory within a DC grid that EMS should connected with each part of the system. Fig.9 represents the architecture of basic DC grid and further EMS of DC grid is classified into two categories – a) classical methods b) artificial intelligence method represented in Fig.7.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:09:41 UTC from IEEE Xplore. Restrictions apply.

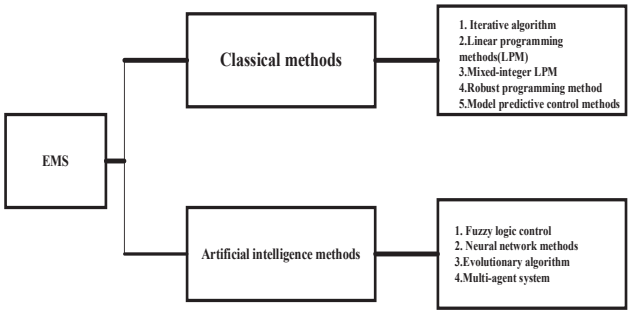


Fig.7 classification of energy management strategy

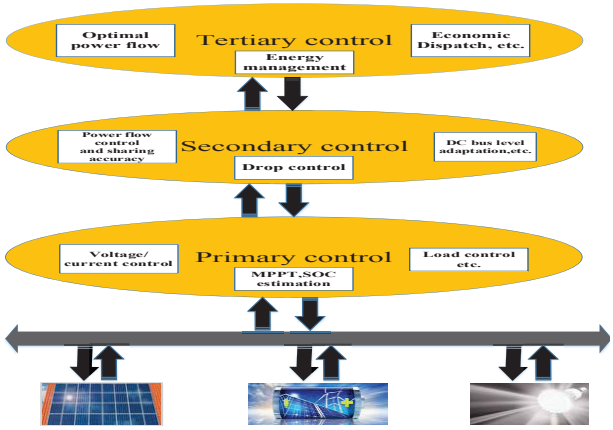


Fig.8 Typical control objectives and function in a DC grid

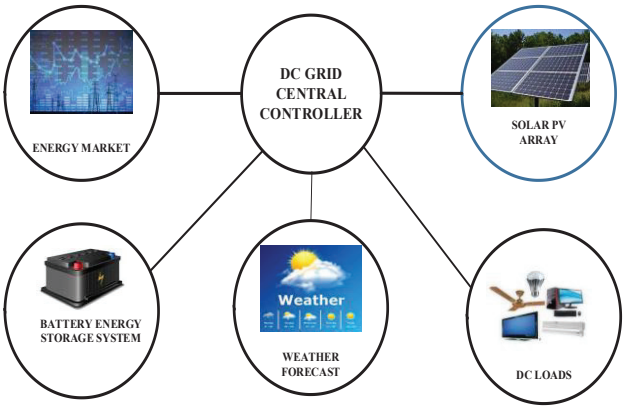


Fig.9 EMS based structure of DC grid

B. *Protection of DC microgrid*

Since DC distributions lacks natural zero points, its protection is a biggest challenge. The 3-phase AC circuit breakers (CBs) were suggested when three contacts gets connected in series. Moreover, the DC protection is used to eliminate the spark. The 3 phase AC circuit breakers is proposed for removing the spark by connecting all the three contacts in series with a DC short circuit protection. With the help of CBs and fuses, the loads and batteries are protected. Moreover, for the protection of DC grid, two types of protections are classified i.e. unit type and non-type. In case of unit type protection scheme it is based upon the information and measurements at both the ends. In case of non-unit protection the system is based upon the information and measured voltages at local ends. So protection scheme is implemented here to provide protection in the particular fixed zones of DC grid, DC bus, converters, battery energy storage system and loads etc. It have the advantage of providing protection coverage of larger area and free operation for protection devices for neighboring zone. The non-unit type protection scheme depends upon threshold setting of electrical quantities. The non-unit protection types also ensures the reliability and necessary discrimination for the DC grid. DC protection scheme evaluated on the speed, selectivity, sensibility and reliability.

V. RESULTS AND DISCUSSION

In order to substantiate of DC source with the household appliances, model has been validated under Simulink platform with the various circumstances. The parameters of system are represented in TABLE I and II. Three level of voltages has been considered 24V (represents low voltage residential DC appliances), 48V (represents medium voltage DC appliances) and 230V (for residential appliances working on AC). The 230V AC bus is obtained from 48V DC bus voltage using high gain converter cascade with SPWM inverter. The 24V DC bus is obtained from the 48V DC bus using DC-DC buck converter. At each voltage level of DC bus, the load resistance have been considered with different switching time. With the case of 48V DC bus, two household load TV (100W) and Laptop (50W) has considered and load value is calculated as 15.36 ohms. The switching time is t= 2secs. Similarly, for the 24V DC bus, two household loads i.e. bulb (100W) and Fan (120W) has considered and the load value is calculated as 2.61 ohms with switching time t=2.5 secs. In case of 230V AC, the load is represented as 26.45 ohms.

Fig.8-10 shows the variation of DC voltage of 48V, 24 V bus and AC voltage of 230V with time. Fig.11.shows the Variation of battery parameters with time. Hence it is observed that the DC voltage at level (48V, 24V) are maintained constant by the proposed technique independent of the amount of solar irradiance, load variation with varying switching time. In other case AC voltage of 230V is able to supply voltages to the residential appliances.

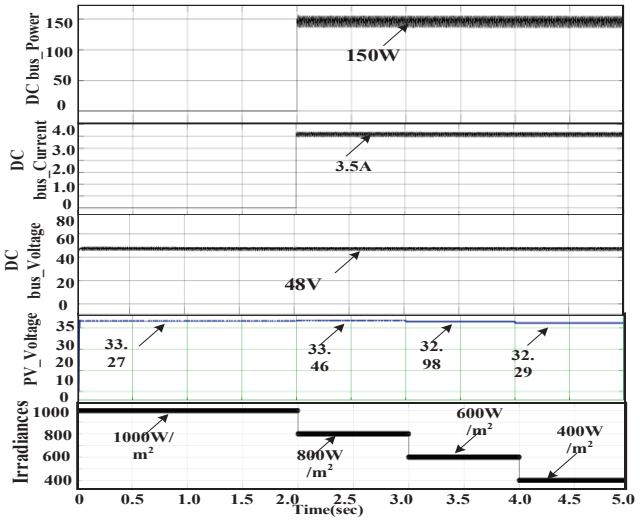


Fig.8.variation of 48V DC bus voltage with time

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:09:41 UTC from IEEE Xplore. Restrictions apply.

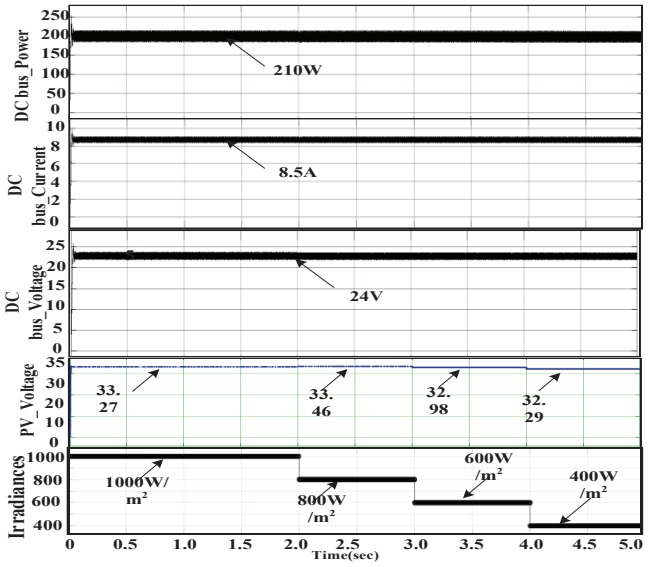


Fig.9 Variation of 24V DC voltage with time

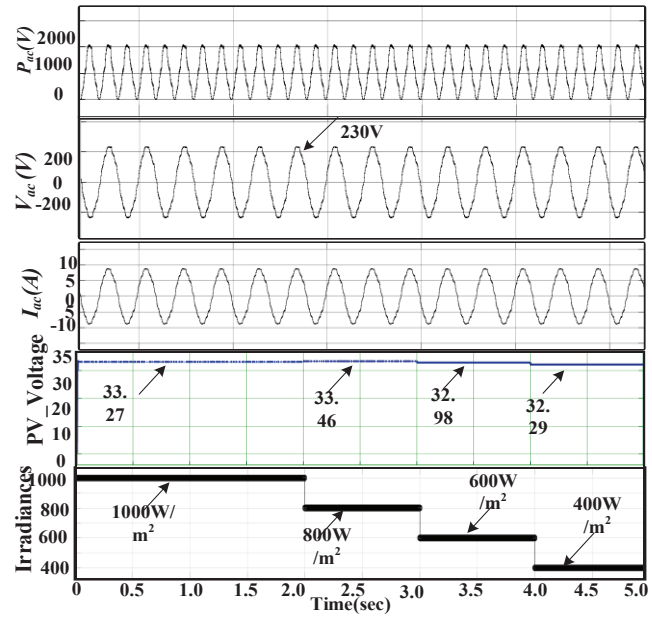


Fig.10 variation of 230V AC bus voltage with time

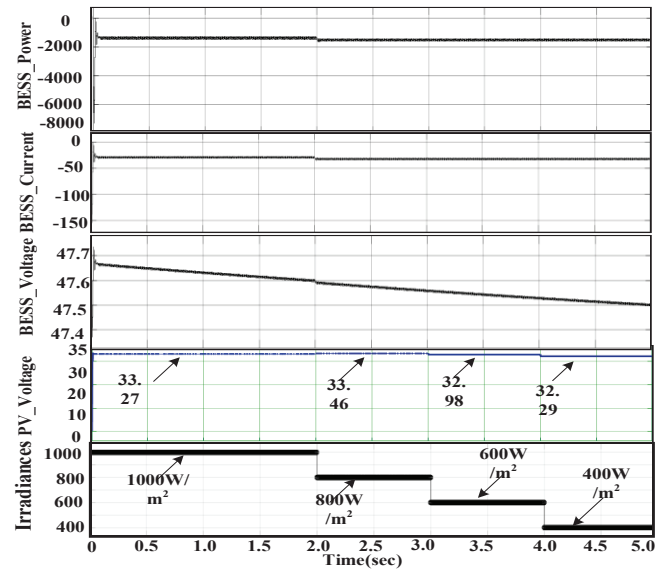


Fig.11. Variation of Battery parameters with time

VI. CONCLUSIONS

On the basis of simulations and modelling, we analyzed the possible ways to provide the DC and AC environment to all necessary homes/offices appliances. The above proposed technique allow the use of DC sources directly from onsite (battery storage + SPV array). It will easily avoid the losses of power due to conversion of (DC →AC→DC). The above model proposes the concept of integration of solar PV array, standalone DC grid and battery energy storage device using bidirectional converter. For MPPT, perturb and observe techniques has been used for the extraction of maximum power from solar panels. From solar PV array, power is supplied to various appliances according to the way of their power consumption ability. The DC bus voltage of 24V and AC voltage 230V is drawn from existing 48V DC bus with using respective converters which is mentioned in section V. The voltage at all the three level maintained constant using battery charging/discharging technique vis-à-vis varying solar irradiance level and load variation. The efficiency of domestic appliances will be much better with the combination of DC voltage of 48V, 24V and AC voltage of 230V for most appliances uses in household applications. The overall power of residential premises is fulfilled by SPV array and battery energy storage system. With decreasing the number of power conversion in the system the reliability and efficiency of distribution system will improve significantly.

REFERENCES

[1] H. K. Mahore and V. Dake, "Modelling and analysis of grid-integrated solar system with power management scheme using battery," *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, 2019, pp. 1620-1623.

[2] A. Khamis, A. Mohamed, H. Shareef, A. Ayob and M. S. M. Aras, "Modelling and Simulation of a Single Phase Grid Connected Using Photovoltaic and Battery Based Power Generation," *2013 European Modelling Symposium*, 2013, pp. 391-395.

[3] M. Kumar, "Solar PV Based DC Microgrid under Partial Shading Condition with Battery-Part 2: Energy Management System," *2018 8th IEEE India International Conference on Power Electronics (IICPE)*, 2018, pp. 1-6.

[4] D. Singh, A. Agrawal and R. Gupta, "Power Management In Solar PV Fed Microgrid System With Battery Support," *2017 14th IEEE India Council International Conference (INDICON)*, 2017, pp. 1-6

[5] A. S. Anees, "Grid integration of renewable energy sources: Challenges, issues and possible solutions," *2012 IEEE 5th India International Conference on Power Electronics (IICPE)*, 2012, pp. 1-6.

[6] S. Park *et al*., "Hardware implementation of optimization technique based sensorless MPPT method for grid-connected PV generation system," *2009 International Conference on Electrical Machines and Systems*, 2009, pp. 1-6 .

[7] P. Gayathri, B. Ashok Kumar and S. Senthilrani, "Control of DC Link Voltage of Single Phase Grid Connected Solar PV System," *2018 National Power Engineering Conference (NPEC)*, 2018, pp. 1-7.

[8] A. Fethi, B. Douadi, M. Laour and B. El-madjid, "Power control of three phase single stage grid connected photovoltaic system," *2016 International Renewable and Sustainable Energy Conference (IRSEC)*, 2016, pp. 1136-1141.

[9] N. C. Sahoo, S. Mohapatro and M. K. Senapati, "A SoC based voltage control strategy for DC microgrid," *2015 IEEE Electrical Power and Energy Conference (EPEC)*, 2015, pp. 185-190,

[10] M. Abu-Zaher, Y. Atia, F. K. Abo-Elyousr and E. El-Zohri, "Experimental Realization for P&O Maximum Power Point Tracking Applied for Single-Stage Three-Phase Grid-Connected Photovoltaic System," *2019 21st International Middle East Power Systems Conference (MEPCON)*, 2019, pp. 704-709

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:09:41 UTC from IEEE Xplore. Restrictions apply.

[11] L. M. Satapathy, A. Harshita, M. Saif, P. K. Dalai and S. Jena, "Comparative analysis of boost and buck-boost converter in photovoltaic power system under varying irradiance using MPPT," *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 2018, pp. 1828-1833.

[12] A. Merabet and R. K. Dhar, "Solar Photovoltaic Microgrid Simulation Platform for Energy Management Testing," *2019 Algerian Large Electrical Network Conference (CAGRE)*, 2019, pp. 1-5

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on October 31,2022 at 09:09:41 UTC from IEEE Xplore. Restrictions apply.

ACS OMEGA

http://pubs.acs.org/journal/acsodf

Review

Functionalized Peptide-Based Nanoparticles for Targeted Cancer Nanotherapeutics: A State-of-the-Art Review

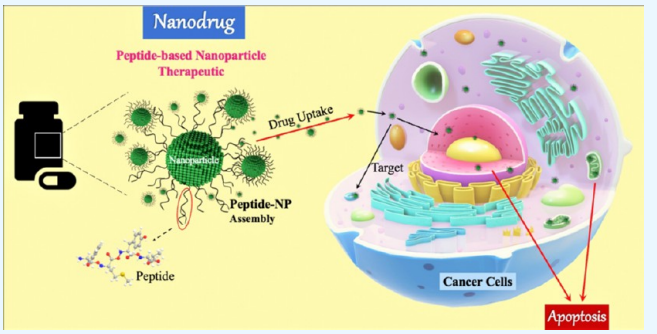
Ritika Sharma, Shikha Jyoti Borah, Bhawna, Sanjeev Kumar, Akanksha Gupta,* Poonam Singh, Vijay Kumar Goel,* Ravinder Kumar, and Vinod Kumar*

Cite This: https://doi.org/10.1021/acsomega.2c03974

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Cancer mortality is increasing at an alarming rate across the globe. Albeit, many therapeutics are available commercially, they are not effective and have no cure up to today. Moreover, the knowledge gap in cancer therapy persists, representing a potential blind spot for the innovation of effective anticancer therapeutics. This review presents an update on current advancements in nanopeptide therapeutics. Herein, a detailed exploration of peptide-functionalized nanoparticles for the development of nanotherapeutics was carried out. Different approaches that include self-assembly nanostructures, solid phase peptide synthesis, ligand exchange, chemical reduction, and conjugation methods for assembling peptides for functionalizing nanodrugs are also highlighted. An outlook on biomedical applications is also reviewed. Additionally, a comprehensive discussion on targeted cancer cell therapy and mechanism of action are provided. The present review reflects the functional novelty of nanodrugs to improve stability, accessibility, bioavailability, and specificity toward cancerous cells. Finally, it summarizes the current challenges and future perspectives on the formulation of these nanodrugs.



Downloaded via ESS INFLIBNET PCA 4 on October 13, 2022 at 06:19:52 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

1. INTRODUCTION

Cancer is a global health issue with limited combinatorial treatment possibilities such as chemotherapy and radiation complemented by adverse effects.[1] However, there is a lack of understanding and clinical translation in cancer pathophysiology. Despite technological advancements and breakthrough achievements in cancer research, there is no cure for cancer. Furthermore, there are certain limitations associated with conventional treatment strategies.[2] Accumulated literature throughout the past decade has provided cardinal insights into the regulatory mechanisms underlying cancer biology and malignancy.[3] Peptides embody a short stretch of amino acids linked by an amide bond (−CONH−) and are involved in various biological functions.[1,4] These biomolecules play a pivotal role in the human body through various biological processes and form an intriguing biological material due to multiple properties such as biocompatibility, bioactivity, and biodegradability.[5,6] The structure−function relationship of peptides have offered several platforms for futuristic breakthroughs in the realm of biomedical applications for various disorders, which has piqued researchers' interest.[5−7] Additionally, therapeutics based on peptides have been of immense interest to researchers due to several features such as low toxicity, better penetration, tunability, and many more.[4] However, there are certain demerits associated with peptide drugs, such as being less stable and prone to degradation.[1,8,9] The concept of supramolecular chemistry and enhanced peptide-based therapeutics through the formation of interactions such as ionic,[10] hydrogen bonding (H-bonding),[11] and even $\pi-\pi$ interactions,[12] results in a systematic hierarchical nanostructure. Despite the lack of understanding of the peptide-nanoparticle (NP) interaction and the parameters that determine its effectiveness, intense research has persisted in achieving ground-breaking results in anticancer medications.

Cancer, being a worldwide incurable terminal disease, creates a global demand for the creation of new materials that show features such as biocompatibility, minimum toxicity, bioavailability, etc. The universe of NPs provides a wide range of materials, among which peptide-functionalized NPs (peptide-f-NPs) has been one of the ideal biomolecules that enhance the efficiency of drug delivery,[13,14] biosensors,[15] COVID-19 vaccines,[16] biotemplated NP catalysts,[17] active ingredients of personal care items (toothpaste and hair

Received: June 25, 2022
Accepted: September 19, 2022

© XXXX The Authors. Published by American Chemical Society

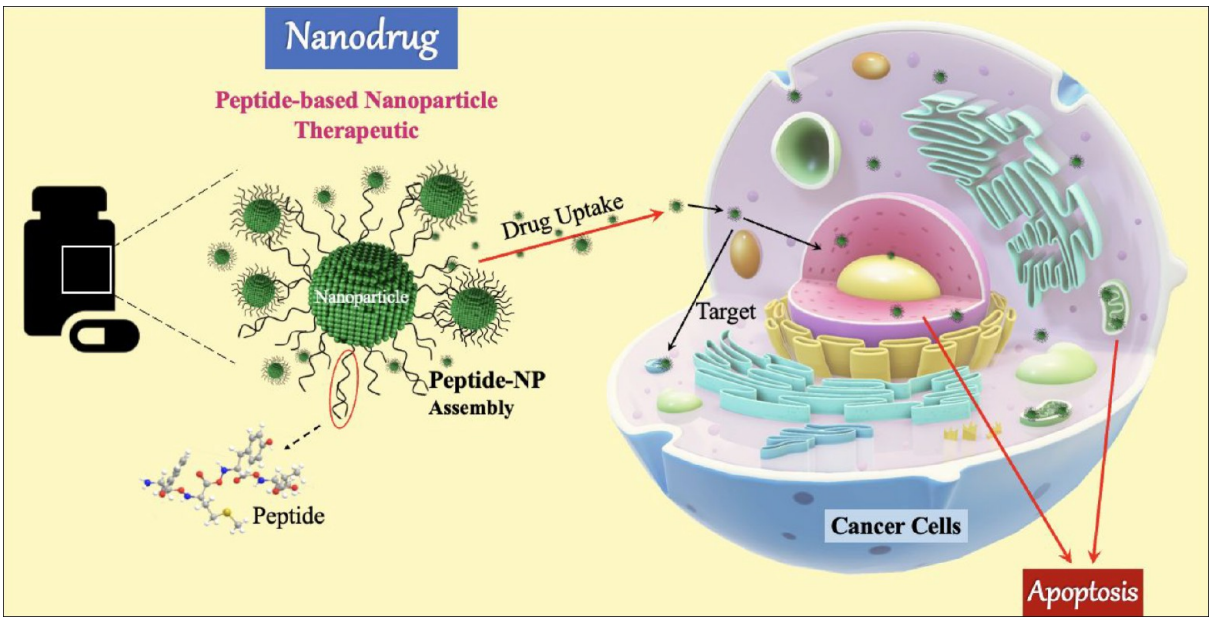


Figure 1. Schematic representation of the peptide-functionalized nanodrug mediating targeted anticancer activity.

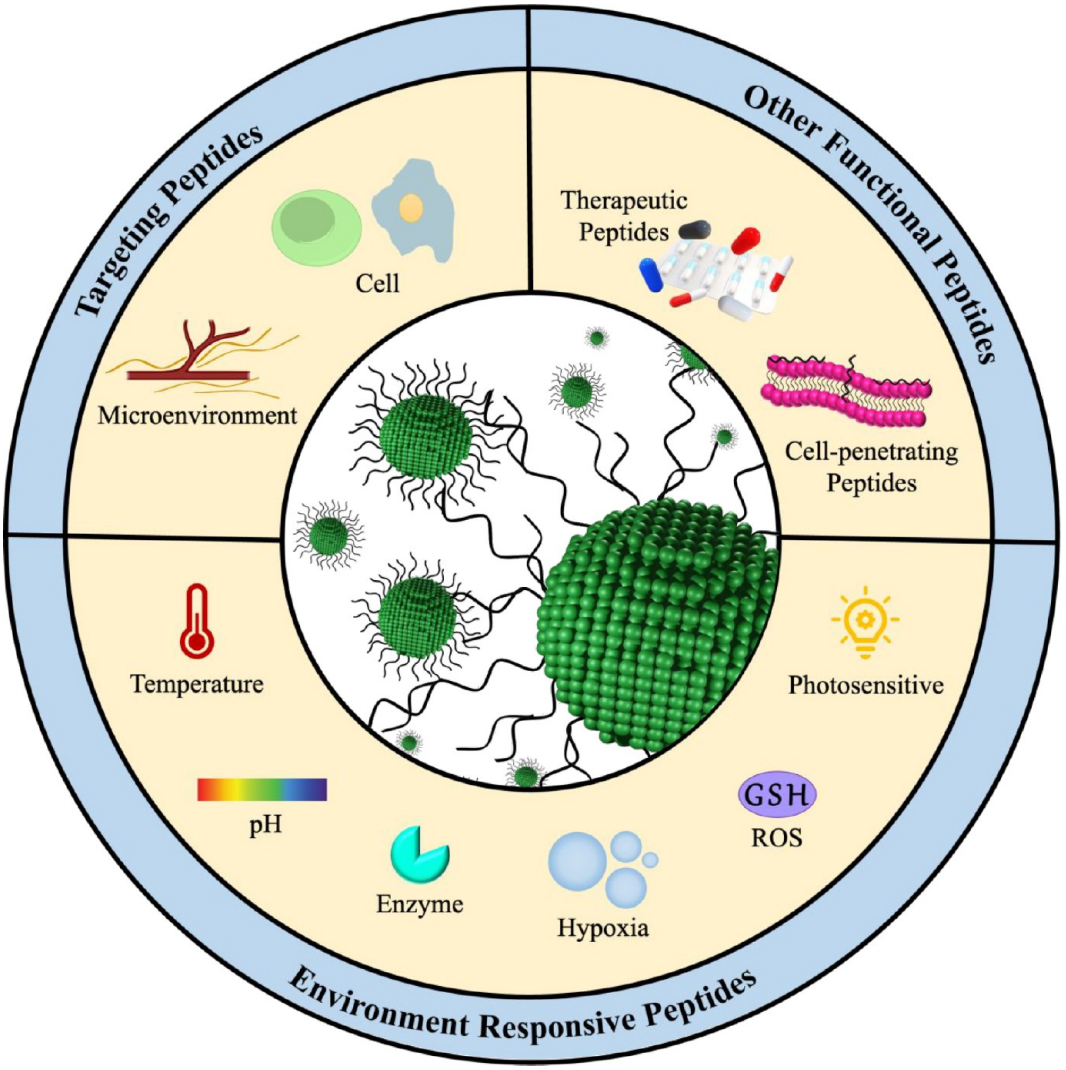


Figure 2. Schematic illustration of peptide-functionalized nanotherapeutics classified into three categories: (1) environment responsive peptides, (2) targeting peptides, and (3) other functional peptides.[23]

https://doi.org/10.1021/acsomega.2c03974
ACS Omega XXXX, XXX, XXX−XXX

coloring),[18] gene therapy,[19] and many other biomedical applications.[20] Consequently, the diverse implementation has urged researchers to explore the *in situ* construction of *in vivo* self-assembly that can be used to develop peptide-f-NPs and determine the possible areas of improvisation.

Therapeutics' based on the combination of peptides and NPs have emerged as a replacement for cancer therapy as they demonstrate the potential for targeted cancer treatment and tracking (Figure 1). Various NPs have been conjugated or functionalized with a peptide moiety via several approaches. The mechanism of targeted cancer treatment is as follows with reduced side effects on healthy cells.[2,21] The peptide-NP assembled therapeutics were injected. Upon injection, the nanopeptides migrate to the tumor cells via penetration or endocytosis. It causes cellular or DNA damage leading to the onset of apoptotic pathways, for example, mitochondria-targeted apoptosis.

From the view of drug discovery and delivery mechanisms, very limited therapeutics are clinically approved.[3] Most of those were not specific to cancerous cells. Moreover, only limited literature explores the functionalization of peptide nanotherapeutics for potential application in targeted cancer therapy.[22] A significant emphasis was provided on the functionalization of NPs. They are summarized with an up-to-date view of anticancer therapy.[23,24] Furthermore, the functionalization of self-assembled nanostructure peptides was also discussed.[25] This work reflects a viewpoint on the functional novelty of nanodrugs with improved stability, accessibility, and bioavailability with cancerous cell specificity.[26] Moreover, the mechanism of action of these nanotherapeutics was also comprehensively discussed. The current challenges and future trends in the formulation of these nanodrugs, along with the future trends are also explored.[27]

2. PEPTIDE ASSEMBLY APPROACHES FOR NANOPARTICLE FUNCTIONALIZATION

Before understanding the application of peptide-f-NPs, an exquisite focus must be brought to their synthesis. Most often, the ideal approach toward the synthesis of NPs is through hydrothermal or solvothermal methods.[28,29] The functionalization of prepared NPs to obtain appropriate efficiency as therapeutic peptide-f-NP drugs is through different strategies: (1) self-assembled nanostructures, (2) solid phase peptide synthesis (SPPS), (3) ligand exchange, (4) chemical reduction, and (5) chemical conjugation for peptide assembly.

2.1. Self-Assembly. For an efficient drug delivery system, a smart binding peptide is required, which can recognize and trigger targeted drug accumulation and release drugs upon appropriate stimuli. Various functional peptides have been determined and are further distinguished based on their functionalities (Figure 2).[23] For instance, targeting peptides can bind to different tissues and lesions (bones, muscles, tumors, etc.), as a consequence of which they can effectively bind to receptors. Examples of such targeting peptides are muscle-targeting peptides (Ala-Ser-Ser-Leu-Asn-Ile-Ala; ASSLNIA) for heart and skeletal muscle diseases,[30] kidney-targeting peptides,[31] and liver targeting peptides.[32] Another class of functional peptides is the responsive peptides, which are produced as a result of changes in pH, light, enzyme, and temperature showing enhanced impact and scope in several applications such as biosensors, tissue engineering, and drug delivery. Zhang et al. synthesized pH and temperature-dependent peptide series, which exhibited reversible transitions between α-helix and β-sheets.[33] In addition to targeting and responsive peptides, therapeutic peptides are the major type of functional peptides, which have presented a promising field for the invention of the drug delivery approach. Many therapeutic peptides have been rigorously studied through clinical trials, besides which many of such peptides have also been Food and Drug Administration (FDA) approved.[34] Having an extensive range of applications in diseases such as cancer, allergies, and metabolic and immunological diseases, peptide-f-NP therapeutics have gained international importance in the global market.[35]

2.2. Solid Phase Peptide Synthesis (SPPS). SPPS is an effectively explored approach, developed by Merrifield.[36] The iterative cycles of deprotection, activation, and coupling are performed using different protecting agents bound to the resin molecule (Figure 3).[24] A frequently used protecting agent is

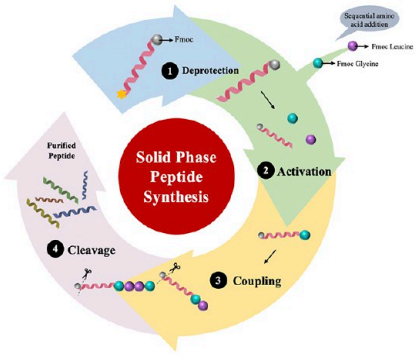


Figure 3. Schematic representation of repeated cycles of Fmoc-based SPPS.[24]

the 9-fluorenylmethoxycarbonyl (Fmoc) group and the *tert*-butyloxycarbonyl (Boc) group, to avoid an undesired reaction. Typically, the Fmoc bound side chain attaches to the resin. Next, the Fmoc deprotection is achieved. Then during activation, amino acids are added sequentially. Subsequently, derivatives of Fmoc-amino acid are hitched using suitable coupling reagents. This coupling step is followed by subsequent deprotection, cleavage, and finally peptide purification.[24,36]

2.3. Ligand Exchange. This simple method proceeds through the displacement of an original present ligand on the surface of the NP by one or a mixture of specifically chosen peptide ligands and has often been seen as a convenient method.[37,38] Ligand exchange occurs under mild conditions of ligands exhibiting multifold functionality and excellent colloidal stability. This displacement has often been successfully observed in peptides having cysteine (Cys; C). For instance, Hu et al. observed that the thiol group can quickly form a stable covalent bond with the gold (Au) atoms on the surface of Au NPs. However, these Au−S covalently bonded interactions may further undergo decomposition by thiols like glutathione, as a result of which modification of Cys using selenium (Se) to produce Au−Se bond formation has been explored by many researchers.[39,40] Alternatively, in mild conditions, phosphorylated amino acids can form stable coordinate bonds when in combination with metallic cations such as Fe^{3+} and Zn^{2+}. The phosphorylated amino acids, like phosphorylated serine (Ser-P), induce an easy exchange of metallic NPs through metal-phosphate coordinate bonds.[41,42] An excellent example of this has been reported by Liu et al.

https://doi.org/10.1021/acsomega.2c03974
ACS Omega XXXX, XXX, XXX−XXX

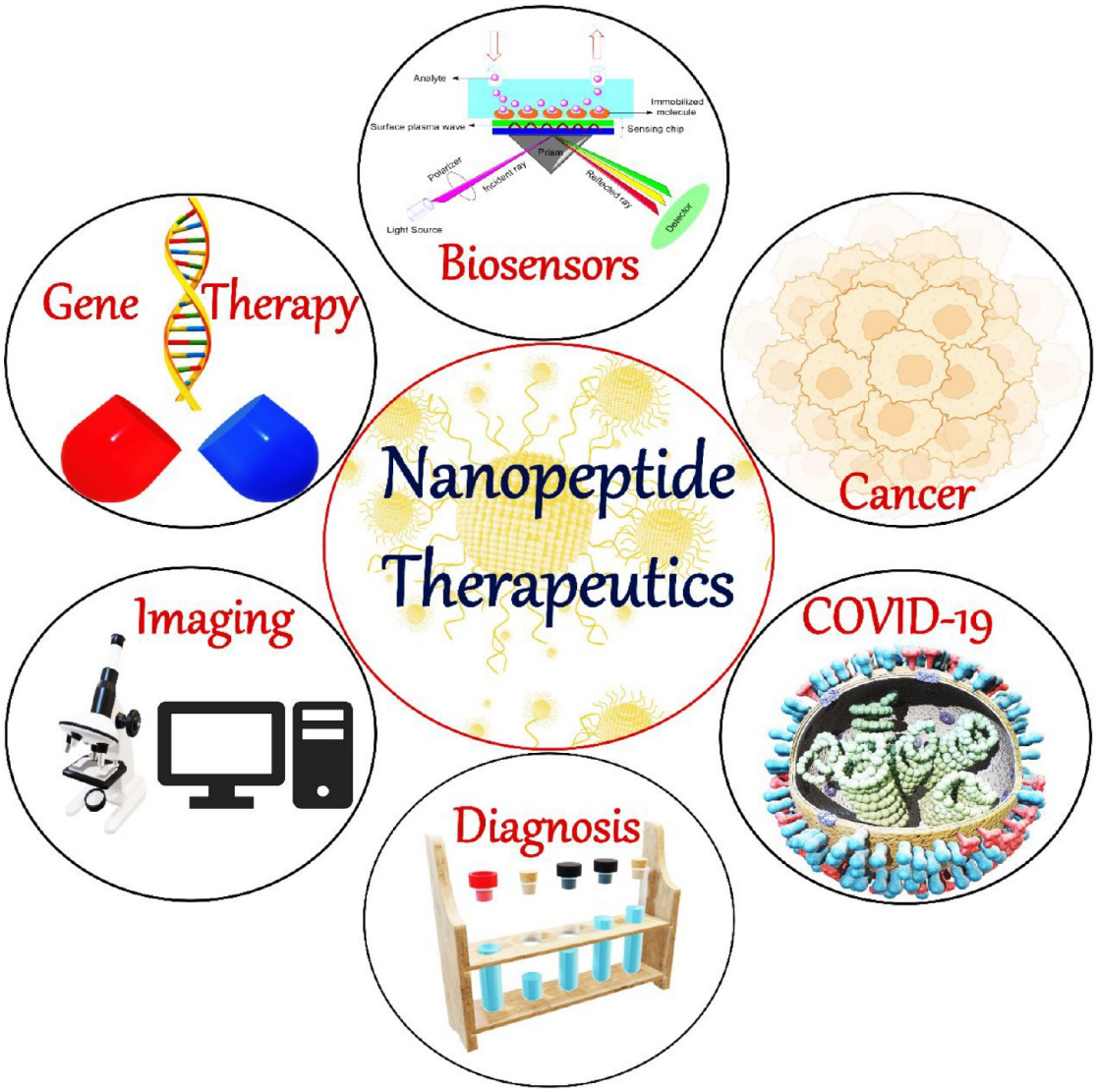


Figure 4. Schematic illustration of nanopeptide therapeutics in biomedical applications.

showing a simple tryptone-induced ligand exchange for achieving functionalization of various NPs such as zinc oxide (ZnO) NPs, silver (Ag) NPs, and $NaGdF_4$ nanodots.[42]

2.4. Chemical Reduction. Chemical reduction is a simple and direct method of synthesizing peptide-f-NPs, which follows three general steps. This method begins with premixing a metal precursor and an appropriate peptide in a solution, followed by the optional second step of adding a reducing agent in a lower quantity. Finally, the synthesized peptide-f-NPs are purified. The peptide incorporated in this method determines the reduction of metal ions as well as the stability of the synthesized NPs.[43] However, the peptide often acts as a stabilizing agent, while several chemical compounds, such as $NaBH_4$ or ascorbic acid, act as the reducing agent. Residual amino acids, like tyrosine (Tyr; Y), tryptophan (Trp; W), proline (Pro; P), and many others, can reduce metal ions to their respective metals via the mechanism of electron transfer.[43,44] This has been seen in recent works by Si and colleagues, who showed a successful reduction of $HAuCl_4$ or $AgNO_3$ via a Trp residue to obtain tripeptide-functionalized Au NPs or Ag NPs.[45] These NPs can be further modified by improvisations based on reaction conditions such as rate of reaction, peptide sequence, metal to peptide ratio, and pH.

2.5. Chemical Conjugation. In the chemical conjugation method, there is a two-step approach to produce desirable peptide-f-NPs. The first step is to cap the NPs with suitable stabilizers such as polyethylene glycol (PEG) derivatives, which proceed via ligand exchange or interactions such as H-bonding, or silica (Si) shells, which proceed via the water-in-oil microemulsion method.[29,46,47] Capping is followed by the conjugation of peptides on the surface of the NPs which can be modulated by different operational parameters. Bartczak and Kanaras successfully conjugated the Lys-Pro-Gln-Pro-Arg-Pro-Leu-Ser (KPQPRPLS) peptide (positively charged) to Au NPs using the N-(3-dimethyl aminopropyl)-N-ethyl carbodiimide hydrochloride (EDC)/*N*-hydroxysuccinimide (NHS) method. Here, the Au NPs were stabilized by carboxy-terminated oligoethylene glycol.[48] Fu et al. developed monocyclic peptides (MCP)-functionalized Mn-doped iron oxide (MnIO NPs) by using a bifunctional PEG-based ligand (3,4-dihydroxy benzylamine-PEG-NH_2) to transfer the MnIO NPs into aqueous solution.[49] A recent work conducted by Li et al. exhibited the fabrication of hydrophobic upconversion NPs (UCNPs) using

https://doi.org/10.1021/acsomega.2c03974
ACS Omega XXXX, XXX, XXX−XXX

a SiO_2 shell from the organic to an aqueous phase and additionally conjugated peptide ligands having an excellent affinity toward tumors. The synthesized peptide-f-NP conjugate was effectively used for diagnosing and bioimaging of colorectal cancer (CRC). Therefore, it holds great significance and has a dual role in cancer therapy.[28]

3. BIOMEDICAL APPLICATIONS OF NANOPEPTIDES

Recent studies have confirmed that peptide-f-NPs have effective potential in antitumor nanomedicines,[50] drug delivery in the central nervous system (CNS)-tuberculosis (TB) treatment,[51] drug delivery,[13,14] nanobiosensors,[15] nanovaccines,[16] gene therapy,[19] cancer, and many other diseases (Figure 4). Regardless, the application diversity has urged researchers to explore other rationales for designing peptide-f-NPs for possible areas of improvisation. A few of them are briefly discussed below. However, the literature already holds a significant number of publications that demonstrated the role of nanopeptides in most applications. Albeit, the literature is scarce on cancer therapy, and in the following section, therefore, an emphasis on the implementation of nanopeptide drugs for focused cancer therapy is elaborated.

3.1. Biosensors and Diagnosis. Multifunctional nanomaterials (MNMs) are the essential foundation for biosensors which are sophisticated tools capable of sensitive and accurate detection. These novel (simple, durable, and sensitive) MNMs used for biosensors have been upgraded, robust, and downsized as nanobiosensors. Recent studies have shown that low molecular weight peptides coupled with nanoscale graphene derivatives are suitable building blocks for analytical formulations with diagnostic capabilities.[52] Peptide-functionalized graphene derivatives have been reported to have multiple biosensing applications such as cancer cell detection,[53] ATP determination,[54] monitoring protease activity,[55] etc. When a targeting peptide is conjugated to the cargo surface for gene delivery, transfection and cell specificity frequently increases. The correct formulations can effectively target *in vivo* gene expression limited to the target tissue and even have beneficial medicinal effects.[56] Li et al. synthesized MoS_2 nanosheets which they functionalized with the Met-Pro-Pro-Pro (MPPP) peptide with the help of polydopamine (PDA) and PEG. The MoS_2-PDA-PEG-MPPP was achieved via facile chemical synthesis. The MoS_2-PDA-PEG-MPPP fluorescent biosensor has been effectively demonstrated for caspase-3 detection with a detection limit of 0.33 ng/mL.[57]

3.2. Imaging. Imaging probes based on nanotechnology are used to quantitatively identify living tissues and cells. Compared with conventional contrast agents, the inclusion of nanoscale materials in the procedure improves the contrast of the interface.[58] In a recent study, Lin et al. developed dual-modal optical imaging rare earth NP (RENP) probes with peptide functionalization for targeted tongue squamous cell carcinoma (TSCC) imaging and therapy.[59] More research has also been explored for applying peptide-f-NPs in targeted drug delivery. For instance, nearly two decades ago, De la Fuente and Berry developed a method to target the cell nucleus by linking the HIV-derived CPP Tat (Gly-Arg-Lys-Lys-Arg-Arg-Gln-Arg-Arg-Arg; GRKKRRQRRR) to Au-NPs. They reported Tat peptides functionalized with Au@tiopronin NPs to demonstrate the capacity to enter cells and target their nuclei.[60] Li et al. used graphene oxide (GO) for the fabrication of a peptide that is specific for gastric receptors, i.e., AF750-6Ahx-Sta-BBN. The biofunctionalization was achieved using H-bonding and $\pi-\pi$ interactions. The affinity for ligand–receptor binding, internalization, and cellular uptake were assessed for HSC-3 cells, which are human oral squamous carcinoma cells. It demonstrated an effectiveness for fluorescent imaging utilizing near-infrared dyes. Additionally, it appears as the potential to offer an NP-based delivery strategy for therapeutic agents.[61]

3.3. COVID-19 Therapy. The most recent global emergency posed by COVID-19 has created a magnificent requirement for advances through the use of nanobased technology and science. This outbreak has put a strain on healthcare and raised important issues regarding their management using traditional strategies and diagnostic devices. In this context, the applications of peptide-f-NPs present a new chance for creating prevention, diagnosis, and treatment solutions against the COVID virus. In a detailed review, Campos et al. have discussed the challenges and drawbacks that need to be addressed when using nanotechnology to manage the COVID-19 virus through the development of nanobased materials, such as disinfectants, personal protective equipment, diagnostic tools, nanocarriers systems, treatment, and vaccine development.[62] Thus, it is clear that peptide-f-NPs have a broad range of applications that have been continuously examined and analyzed over the last three decades. However, further research is necessary to fully comprehend and utilize this potential.

3.4. Gene Therapy. A recent investigation by Tarvirdipour et al. demonstrates the smart assembly of NPs that are functionalized with peptide molecules. They comprehensively presented the potential for the implementation of peptide-f-NPs for their application in gene therapy and tracking. The demonstration of nanobased delivery of nucleic acids thereby enhances the probe sensitivity for targeted therapy.[19] One study demonstrated PEG along with octaarginine (R8) peptide. The Arg-rich peptide molecule is used as a cell-penetrating peptide that can be a potential cargo for delivery to different cells. Their investigation showed pH-responsive behavior of the synthesized PEG-plasmid DNA (pDNA)-R8 peptide as well as synergistic effects.[63] Shtykalova et al. investigated magnetic NPs that are biofunctionalized with the hexaarginine (R6) peptide, which is bound electrostatically with the $\alpha v\beta 3$ integrin targeted peptide. The functionalization was achieved and was conjugated with pDNA. It reduces transfection time and improves delivery of pDNA encoding thymine kinase gene for the herpes simplex virus type I (HSV-1). Therefore, it shows a potential nonviral-based approach for the treatment of uterine leiomyoma.[64]

4. CANCER THERAPY

Nanoscale intervention over the past decades has provided insightful information for peptide-f-NP based on therapeutic approaches for cancer treatment and diagnosis. Several NPs, such as Au, GO, quantum dots (QDs), and others, have been integrated with peptide drugs to enhance the stability of peptides. Furthermore, peptides combine immense potential and large protein selectivity, possessing specific pros of small molecules, like stability, oral bioavailability, and availability. Therefore, various peptide-f-NPs have been developed, such as self-assembly, supramolecules, cell-penetrating peptides, peptide-based nanopores, peptide-based hormones, stimuli-responsive peptides, and layered double hydroxide (LDH)-based peptides.[1,4,8,13,24,65−68] Recent advances in the peptide-f-NP nanodrug would increase their potential applications in

https://doi.org/10.1021/acsomega.2c03974
ACS Omega XXXX, XXX, XXX−XXX

Table 1. Overview of Peptide-f-NP Drugs Used As Anticancer Therapeutics

peptide	f-NP or loaded drug	synthesis method	application in cancer therapy	ref
PS NPs				
CpG; E7; WT1; SV	PS NPs	covalent conjugation	elicit CD8+ T cell specificity with minimal cross-reactivity in gynecological cancers	69
FMYL	PS NPs	GLIDE docking	anticancer agent	21
		facile SPPS	inhibit the DHFR pathway in cancerous cells	
PEG-based Materials				
TAT	PEG−PCL	ring-opening polymerization	encapsulation of siRaf-1 and camptothecin in the formed peptide-NP	70
			exhibit higher uptake and beneficial effect in glioma cells	
MP9-aPDL1	PEG	SPPS	enhanced CRC immunotherapy	71
Au NPs				
CRGDK; P12; Nrp-1	Au NPs	EDC/NHS	improved therapeutic P12 peptide delivery to cancer cells	72
			highly effective breast cancer treatment (upregulating p53 expression)	
LA-WKRAKLAK	Au NPs	facile mixing	induces intrinsic apoptotic pathway in cancer cells	73
Other Materials				
SMAC	DOX	EDC/NHS	treatment of drug-resistant cancers	74
FRRG			highly specific to cancer cell	
D-penetratin	DOX	SPPS	exhibits increased uptake in brain perfusion studies	75
SynB1			enhanced DOX (20-fold) transported into the brain parenchyma	
			reduced DOX concentration in heart tissue	
aPDL1 and cathepsin-B cleavable peptide	DOX	intermolecular interactions	targeted cancer immunotherapy	76
			induced ICD	
RGDK receptor	DOX	EDC/NHS method	efficient delivery of cytostatic drugs	77
			tumor inhibition upon single dose administration	
Fe(III)-DA	DOX	SPPS	efficacy of CDT and PTT	78
mussel-derived peptide			photothermal ablation	
			multimodal synergistic cancer therapy	
P435 HER2/neu-derived	Maleimide-PEG2000-DSPE	covalent binding with a thiol group	nanoliposomal vaccine delivery systems against HER2-positive breast cancer cells	79
streptavidin-coated peptide (TLR9 and CpG ON)	PLGA	double emulsion solvent evaporation	potential as a therapeutic cancer vaccine	80,81
			effective against melanoma or prostate cancer cells	
F-KLAK	poly-CB[7]	facile mixing	∼97% encapsulation without affecting the peptide activity	82
			enhanced stability and anticancer potency of the peptides *in vivo*	
FFYSV	PpIX	facile SPPS	tumor PA imaging	83
			chemo-photothermal combination therapy	
PTX; CPK	PLGA	emulsion solvent evaporation	high affinity for MCT1 receptor	84
			potential antineoplastic nanodrug	
			specific for CRC cancer	
			inhibit angiogenesis	
MiRGD	Graphene QDs	hydrothermal method	multifunctional for targeted drug delivery and tracking	85
	DOX or Curcumin	chemical synthesis		

anticancer activity. Table 1 summarizes the peptide-f-NP drugs along with their application in tumor targeting.

To address this, peptide-f-NPs have been of interest in association with cancer detection, imaging, and therapy due to the aforementioned characteristics. As tabulated above, the recent trends associated with cancer therapy have been classified based on the carrier materials: (1) polystyrene (PS) NPs, (2) PEG-based, (3) Au NPs, and (4) other nanomaterials.

4.1. Polystyrene NPs. Singh et al. have fabricated human dihydrofolate reductase (hDHFR) for the first time. In their study, they performed molecular docking studies for the hDHFR along with its substrate (PDB ID 4M6K) and docked the peptides via the GLIDE docking program. Their results demonstrated a tetrapeptide conjugated with carboxylated PS NPs, enhanced delivery at the tumor site, and enhanced growth inhibitory effects. They designed a tetrapeptide (Phe-Met-Tyr-Leu; FMYL) that can efficiently bind at the substrate binding site of hDHFR, instead of folic acid or methotrexate. Therefore, conjugation of FMYL to PS NPs improves the delivery of the optimized peptide and growth-inhibiting effects. Therefore, this rationally designed peptide-f-NP drug possesses the ability to overcome the limitations of a conventional anticancer agent, i.e., methotrexate.[21] Xiang et al. have formulated a peptide vaccine with the integration of PS NPs. The peptides used are the E7 protein of the human papillomavirus, 5′-cytosine-phosphate-guanine-3′ (CpG), surviving (SV), and Wilms Tumour antigen 1 (WT1). These peptides were covalently conjugated with PS NPs and were tested for anticancer activity. Their results demonstrated that

https://doi.org/10.1021/acsomega.2c03974
ACS Omega XXXX, XXX, XXX−XXX

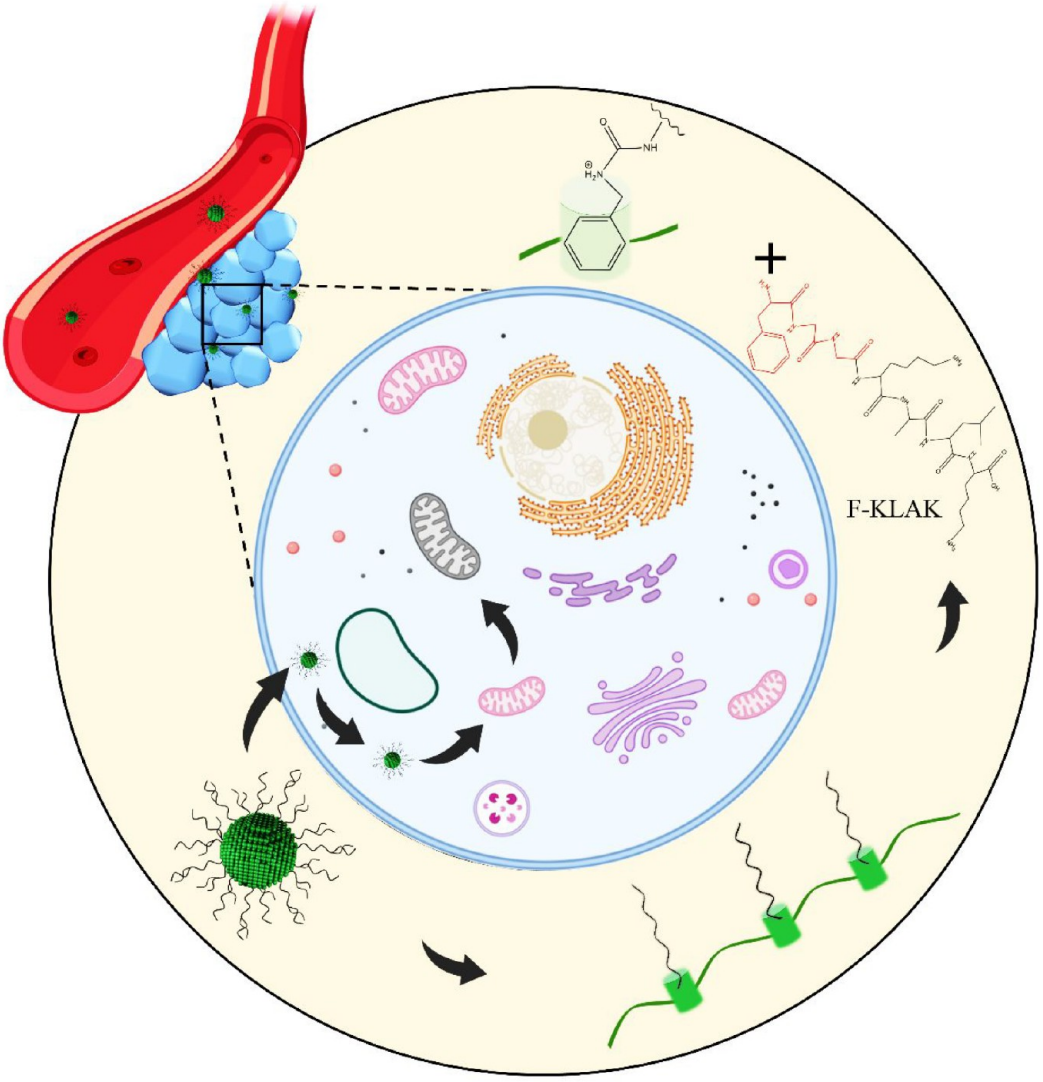


Figure 5. Schematic illustration of the development of supramolecular peptide therapeutic (SPT) for enhanced stability and anticancer ability of the peptides *in vivo*.[82]

the formed nanovaccine elicits a CD8+ T cell response without any inflammation. Therefore, these are effective against gynecological tumor cells.[86]

4.2. PEG-Based Materials. Wang et al. have demonstrated a cucurbituril- and phenylalanine (Phe; F)-functionalized peptide-f-NP drug via the facile mixing method. In the synthesis, curcubit[7]uril (CB[7]) was copolymerized with PEG. Further, the N-terminal F-containing peptide (Phe-Lys-Leu-Ala-Lys; F-KLAK) and poly-CB[7] were mixed in equimolar ratios and were stirred under ambient temperature and pressure. The poly-CB[7]/F-KLAK (PCB/FK) composite was synthesized, and the degree of polymerization was measured through proton nuclear magnetic resonance (^{1}H NMR) spectroscopy. The PCB/FC complexed nanodrug demonstrated ∼97% encapsulation without affecting peptide activity, and the designed nanodrug was studied *in vivo* and their potential role in anticancer efficacy was observed (Figure 5).[82] A recent investigation demonstrated enhanced CRC immunotherapy using PEG conjugation with mastoparan (MP1) and anti-programmed cell death 1-ligand 1 (aPDL1) via SPPS. The prodrug was conjugated using a matrix metalloproteinase-2 (MMP-2) linker. The PEG-MPL1-aPdL1 conjugate exhibited oncolytic and pharmacokinetic properties along with the immune checkpoint block (ICB). The novel peptide-f-NP are accumulated in tumor cells demonstrating target specificity and enhancing infiltration of lymphocytes to induce oncolytic immunotherapy.[71]

4.3. Au NPs. One of the studies by Akrami et al. demonstrated the synthesis of pro-apoptotic peptides, i.e., α-lipoic acid (LA) and peptide conjugate (Leu-Ala-Trp-Lys-Arg-Ala-Lys-Leu-Ala-Lys; LA-WKRAKLAK) which was functionalized with Au NPs via facile mixing. The formed LA-peptide-Au NP conjugate could induce intrinsic apoptosis in anticancer cells. Furthermore, they demonstrated that the performance of the formed conjugate depends on the morphology features, like size and shape, of the NPs. For example, nanospheres showed cytotoxicity and morphological changes over nanorods. Similarly, nanorods demonstrate higher hemolytic activity over nanospheres.[73]

Kumar et al. have functionalized Au NPs along with Cys-Arg-Asp-Gly-Lys (CRDGK) and p12/PMI, that is, targeted and therapeutic peptides, respectively. The Au@P12-CRDGK was synthesized via the EDC/NHS method (Figure 6A,B).[72] On the basis of the available literature, p12 is known to

https://doi.org/10.1021/acsomega.2c03974
ACS Omega XXXX, XXX, XXX−XXX

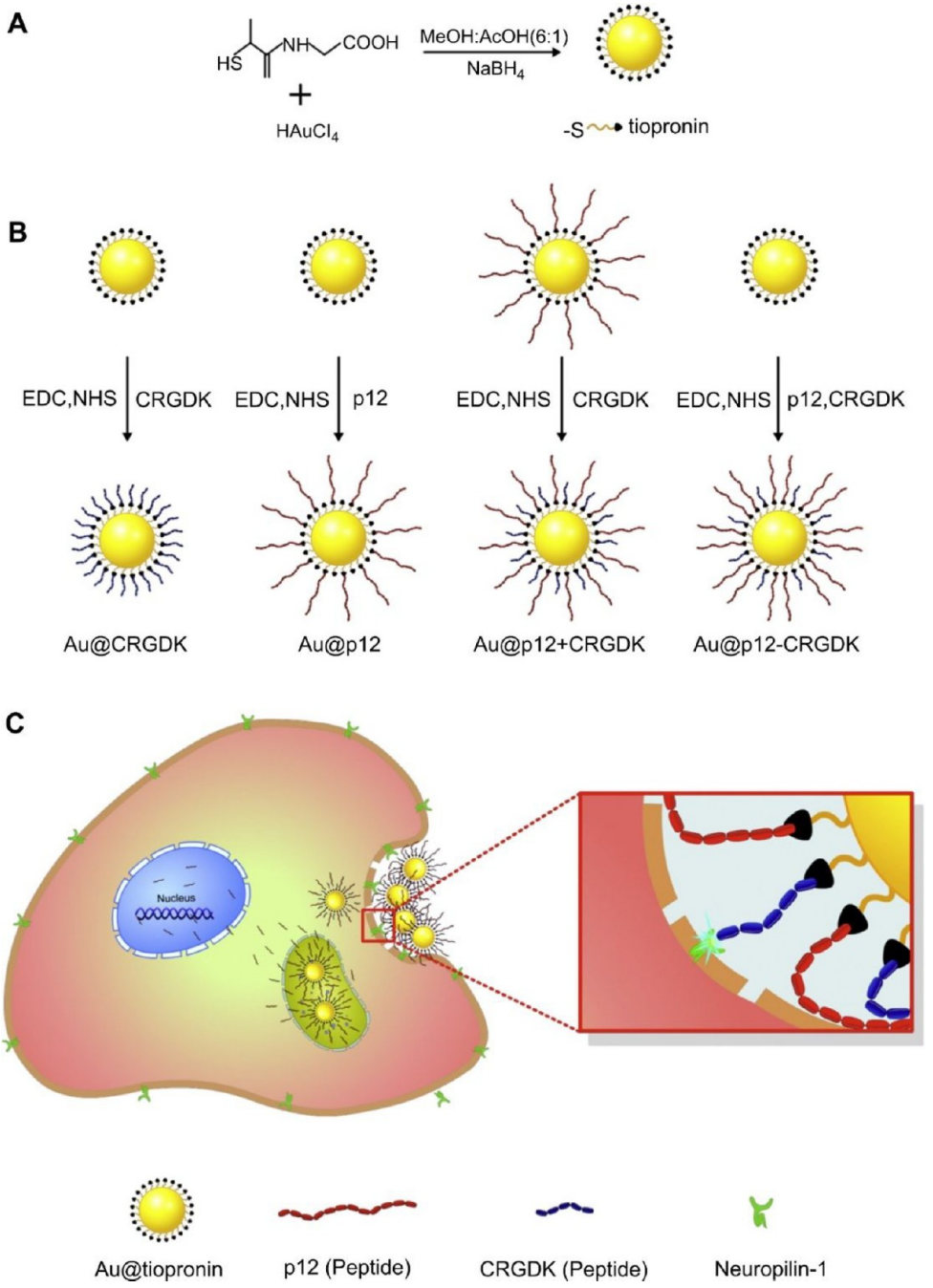


Figure 6. Schematic representation for Au@tiopronin synthesis and peptide functionalization for cancer treatment. (A) Synthesis of Au@tiopronin. (B) Au NPs functionalized with two types of peptides: p12 and CRGDK. (C) Receptor−ligand interaction enhances intracellular entry and increases response to the intracellular release of the therapeutic peptide into the cells.[72] Copyright 2011 Elsevier Ltd.

regulate p53 (a tumor suppressor protein) and is thereby responsible for cell cycle inhibition, DNA damage, and apoptosis. Furthermore, neuropilin (Nrp-1) binds with the targeted peptide that interacts with the protein tyrosine kinase receptor and thereby results in angiogenesis in breast cancer.[87−92] The conjugation of Au NPs enhances the uptake of the therapeutic peptide, and the selective binding of the targeted peptide with the Nrp-1 receptor causes a targeted effect on the cancerous cells without harming healthy neighboring cells (Figure 6C).[72]

4.4. Other Nanomaterials. Shim et al. have fabricated a second mitochondria-derived activator of caspase/Phe-Arg-Arg-Gly/doxorubicin (SMAC/FRRG/DOX) and cathepsin-B (SMAC/Phe-Gly-Phe-Arg (FGFR)/DOX) specific conjugates via the EDC/NHS method and further sieved through reverse-phase high-pressure liquid chromatography (RP-HPLC). The chemical structures for chemical conjugation, demonstrating hydrophobic interaction, and $\pi-\pi$ assembly between the peptide and DOX NPs were developed (Figure 7a).[74] SMAC-FRRG/FRGR-DOX was fabricated, and the drug−drug (D-D) NPs accumulated in cancerous cells via enhanced permeability and retention (EPR) effect, i.e., the accumulation of peptide-f-NP drugs in the tumor cells thereby enhancing permeability (Figure 7b). The NP composite got cleaved, and free DOX causes pro-apoptosis in cancerous cells mediated by cathepsin-B (Figure 7c). Briefly, the formed conjugate specifically cleaves

https://doi.org/10.1021/acsomega.2c03974
ACS Omega XXXX, XXX, XXX−XXX

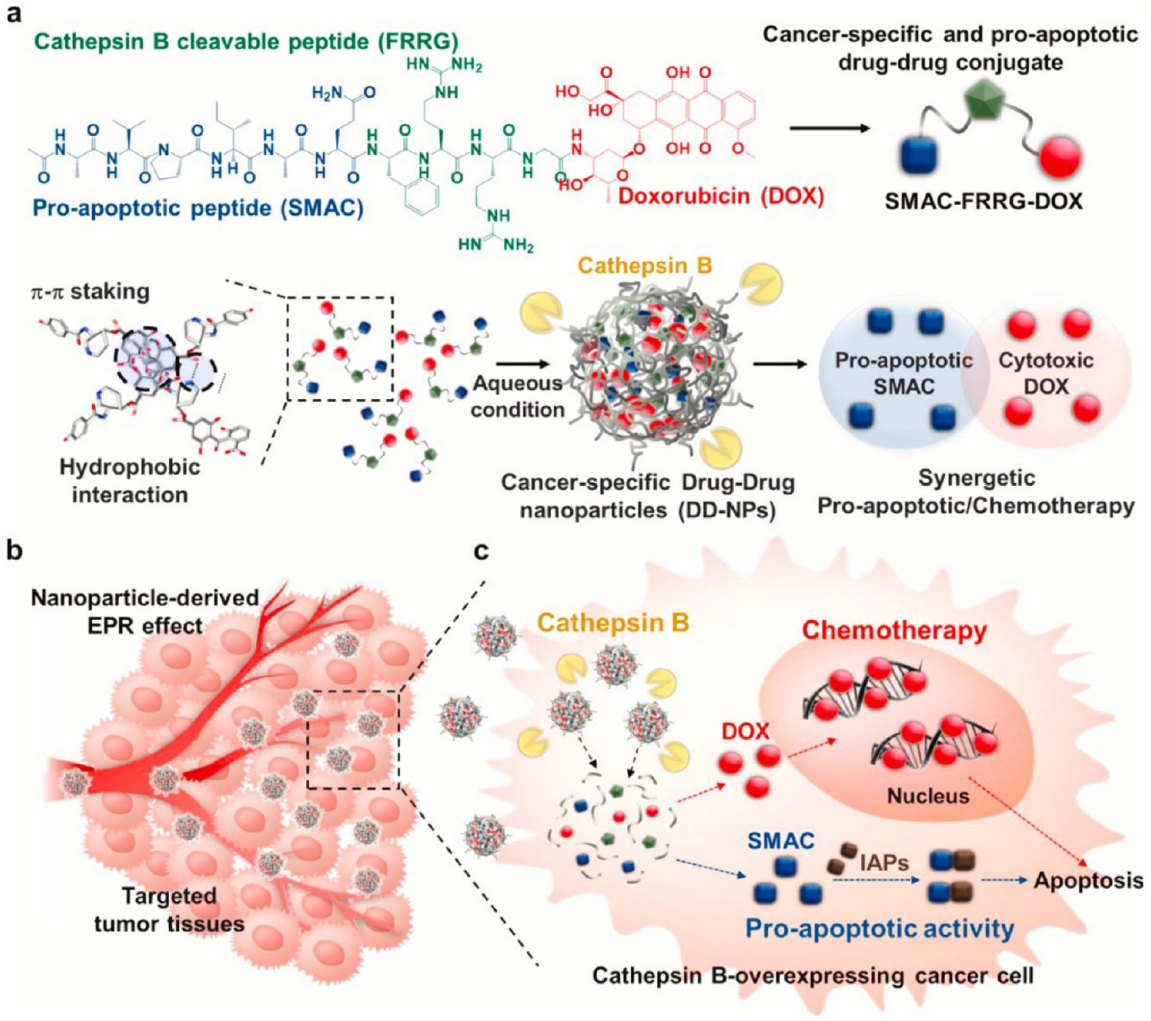


Figure 7. Synergetic pro-apoptotic/chemotherapy by cancer-specific and pro-apoptotic DD-NPs. (a) Chemical structure of cancer-specific and pro-apoptotic D-D conjugate that can form stable NPs. (b) The DD-NPs highly accumulate in the targeted tumor tissues via the EPR effect. (c) The conjugated NPs are specifically cleaved, resulting in drug-resistant cancer cells. Copyright 2020 Elsevier Ltd.[74]

pro-apoptotic SMAC and cytotoxic DOX in cancer cells that are overexpressing cathepsin B, and then pro-apoptotic/chemotherapy is influenced via the synergic effects of the chemotherapy and the inhibitor of apoptosis proteins (IAPs) inhibition. Contrastingly, due to the reduced cytostatic cathepsin-B expression in normal cells, the novel strategy depicted reduced toxicity in normal cells. Therefore, the novel nanotherapy is effective for pro-apoptotic and chemotherapy, with the potential to tame drug resistivity in cancerous cells.[74] Another investigation demonstrated an efficient utilization of D-penetratin or SynB1 in conjugation with DOX via SPPS. The formed conjugate showed a 20-fold enhancement in the uptake to the blood-brain barrier.[75]

Another study explored cathepsin-B cleavable peptide and aPDL1 by conjugating them to DOX through intermolecular interactions. The novel prodrug entered the tumor cells via a receptor (programmed cell death-ligand 1; PDL-1) mediated endocytosis. After entry, DOX was released, which induced immunogenic cell death (ICD). Synchronously, pathways responsible for immune suppression were disrupted, resulting in the infiltration of T lymphocytes. The prodrug particles accumulated in the cancerous cells and resulted in innumerate T cell recruitment.[76]

A recent investigation demonstrated the fabrication of DOX that was stacked with mesoporous Si NPs. Furthermore, the formed nanocarrier was conjugated with an aptamer (AS1411) as well as the Arg-Gly-Asp-Lys (RGDK)-octa-arginine (RGDK receptor) peptide. The formed conjugate was elucidated for efficient delivery of cytostatic drugs (Figure 8). The results elucidated breast cancer tumor inhibition upon administration of a single dose via *in vivo* analysis.[77]

A research group explored DOX-loaded NPs for synthesizing mussel-derived peptides via SPPS. The complex has the ability for self-assembly to form the Fe(III)-dopamine (DA) complex. The complex demonstrates chemodynamic therapy (CDT) and photothermal therapy (PTT). As the pH changes, the DOX is released in the cell (in an acidic environment), inducing apoptosis and increasing ROS through Fenton reactions. Therefore, it has potential for cancer therapy with diminished side effects (Figure 9).[78]

Ren et al. have fabricated porphyrin (PpIX) and Phe-Phe-Tyr-Ser-Val (FFYSV) peptides via SPPS, possessing the ability to self-assemble supramolecular nanodrugs (Figure 10a).[83] The PpIX-FFYSV nanodrug gets accumulated in cancerous tissue. Upon irradiation, PpIX converted light to heat for imaging tumors (Figure 10b). Furthermore, peptides acted as

https://doi.org/10.1021/acsomega.2c03974
ACS Omega XXXX, XXX, XXX−XXX

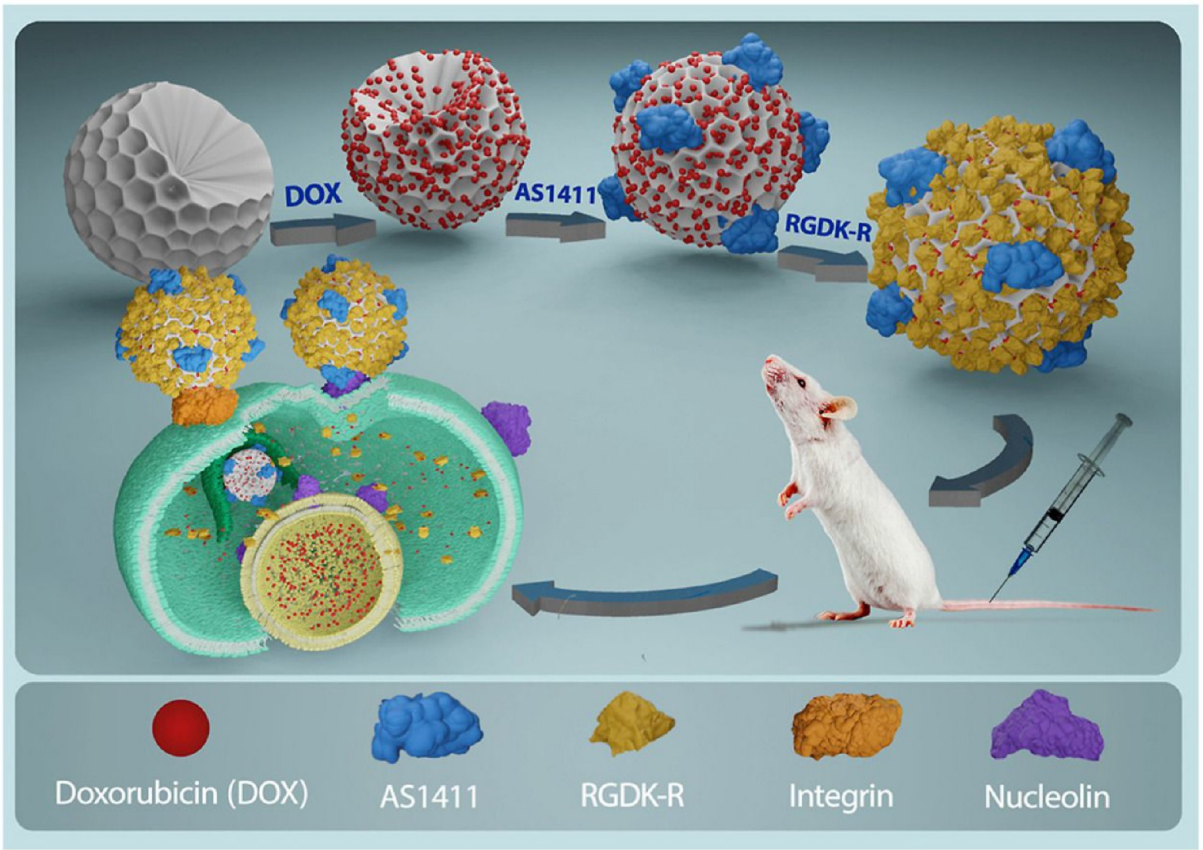


Figure 8. Structure of RGDK receptor-Apt-DOX-MSN.COOH (MDAP) and its function.[77] Copyright 2022 Elsevier B.V.

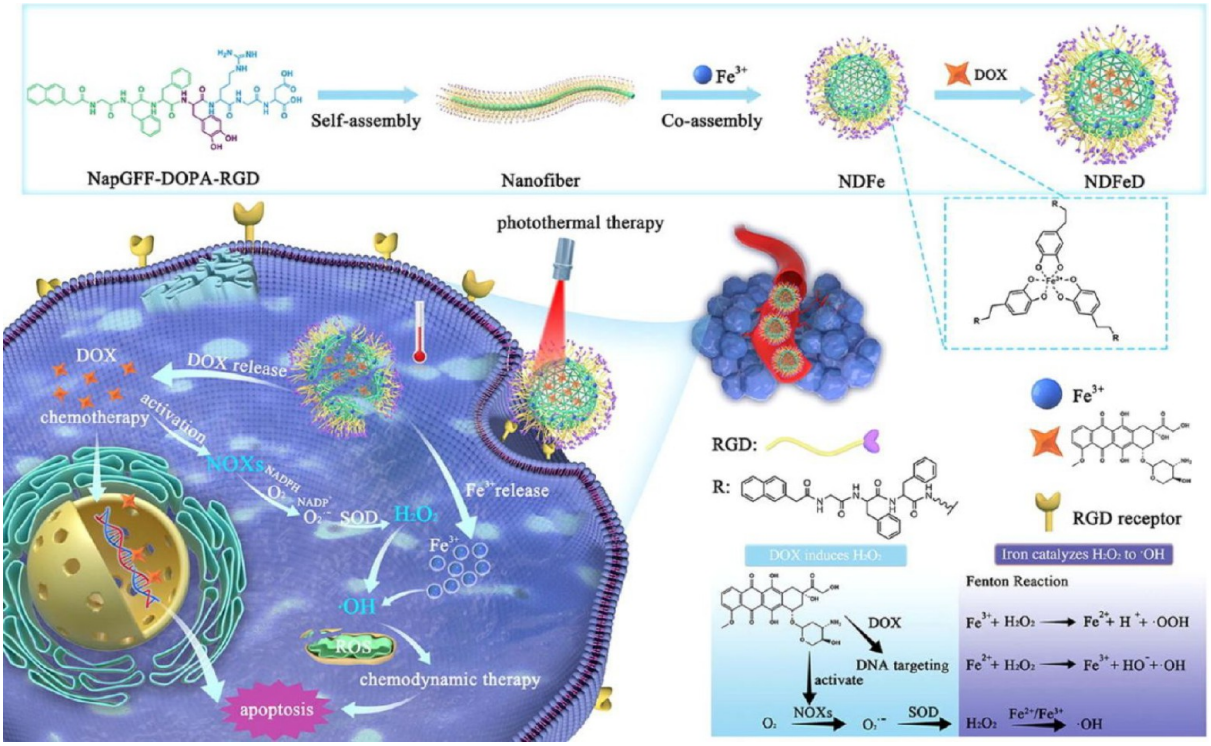


Figure 9. Schematic illustration of the self-assembly process of the mussel-derived peptide (NapGFF-DOPA-RGD), Fe^{3+}, and DOX for the generation of multifunctional NMs with potential for synergistic PTT-CDT-chemotherapy.[78] Copyright 2022 Elsevier B.V.

histone deacetylase inhibitors, which led to apoptosis and increased acetylation of histone. Hence, the designed supra-

https://doi.org/10.1021/acsomega.2c03974
ACS Omega XXXX, XXX, XXX−XXX

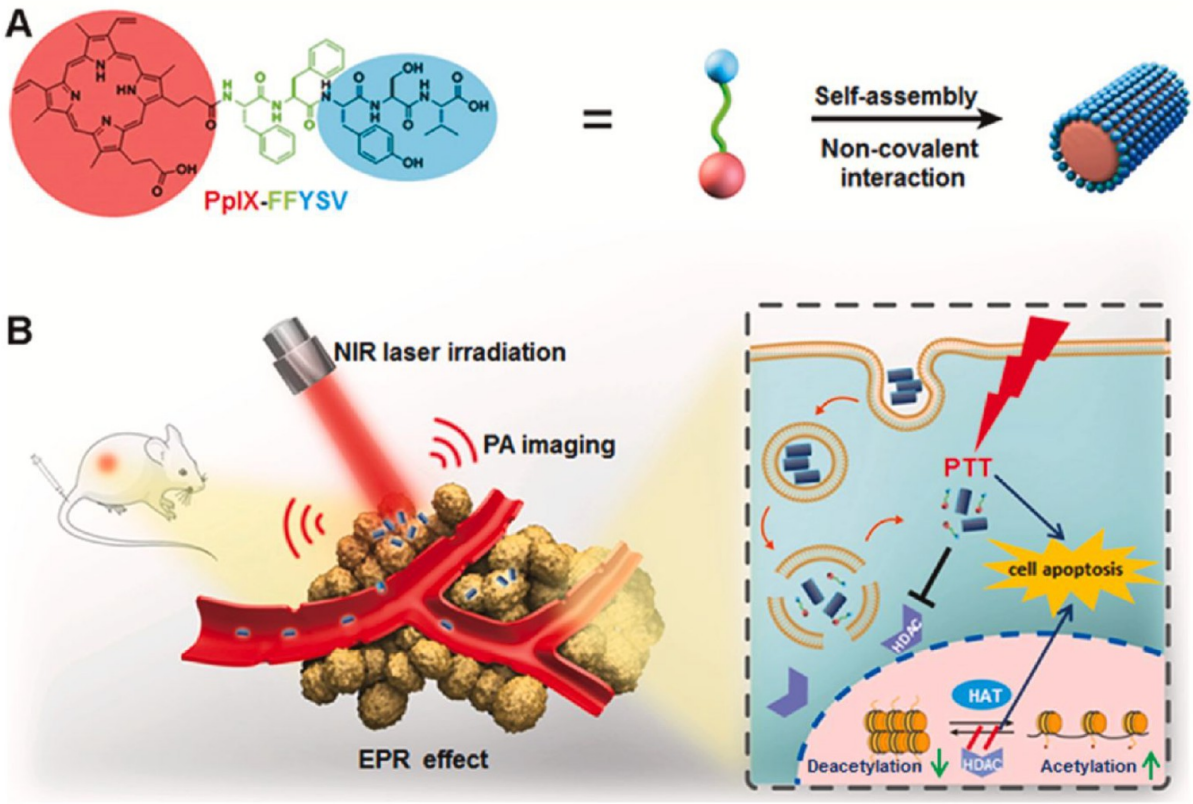


Figure 10. (a) Molecular structure of PpIX-FFYSV and illustration of the monomeric self-assembly into PpIX NAs. (b) Schematic illustration of the self-assembled PpIX NAs for *in vivo* tumor photoacoustic imaging and chemo-photothermal combination therapy.[83] Copyright 2021 Elsevier B.V.

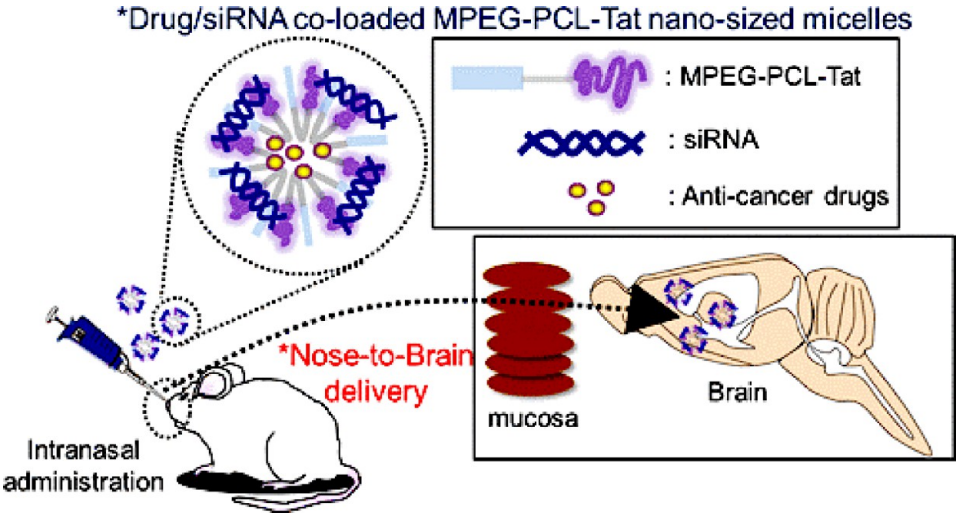


Figure 11. Schematic illustration of intranasal drug/siRNA co-delivery to the brain with cell-penetrating peptide-modified nanomicelles. Reproduced from ref 70. Copyright 2014 American Chemical Society.

molecular nanodrugs with self-assembly ability could be a potential platform for chemo-photothermal nanotherapy and photoacoustic (PA) imaging of cancerous cells.[83]

Kanazawa et al. have fabricated a nanomicelle-based intranasal drug using an altered poly(ethylene glycol)-*co*-poly(ε-caprolactone)-Tat block copolymer (MPEG-PCL-TAT) via a ring-opening polymerization approach. These nanomicelles were demonstrated to be stable with a particle size range between 60 and 200 nm and were found to exhibit potential application as an intranasal therapeutic for delivering the drug to the brain and CNS (Figure 11).[70]

Farzad et al. have fabricated the P435 HER2/neu-derived (IRGRILHDGAYSLTLQGLGIHGGGC) peptide that was linked to maleimide-PEG2000-DSPE (1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine) via an adjuvant: 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE) and monophosphoryl lipase A and covalent linkage between the conjugate and thiol group. The formed conjugate of the nanoliposomal

https://doi.org/10.1021/acsomega.2c03974
ACS Omega XXXX, XXX, XXX−XXX

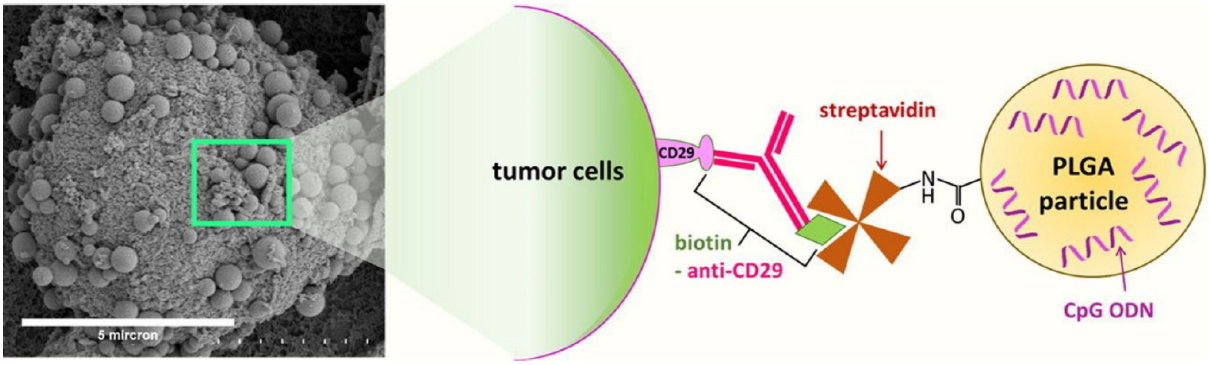


Figure 12. Schematic representation of cell-surface engineered particle with potential as a therapeutic cancer vaccine.[81] Copyright 2017 Elsevier B.V.

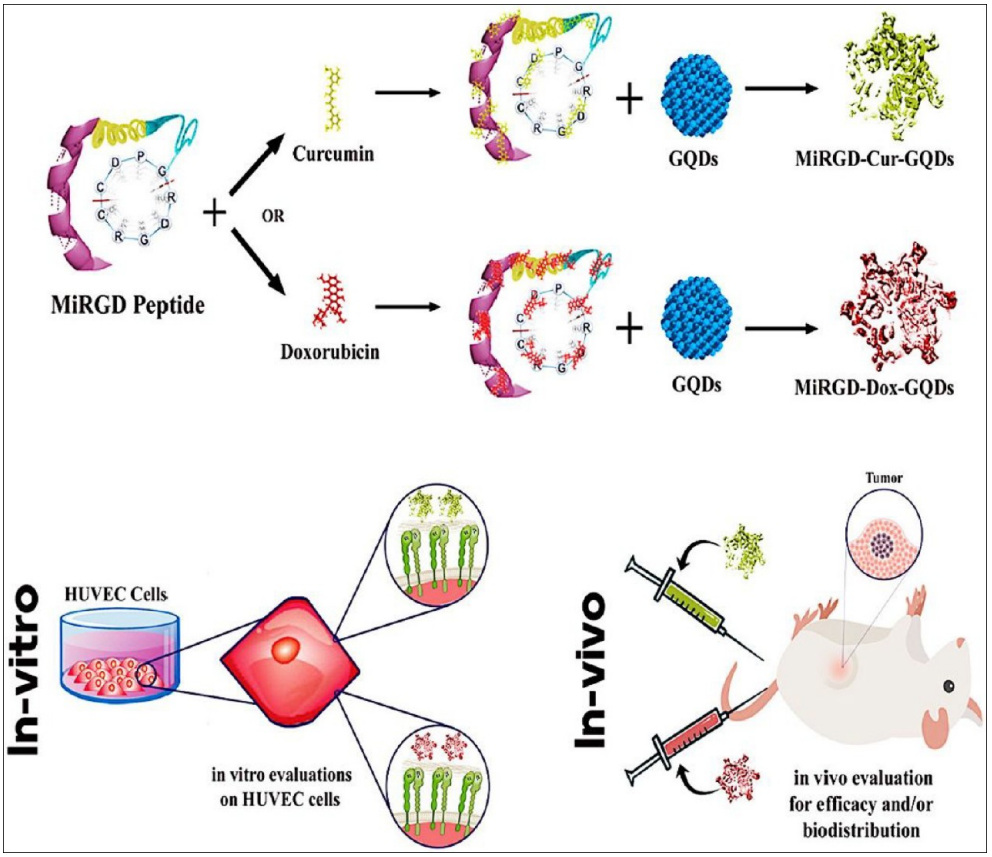


Figure 13. Schematic illustration of peptide-functionalized with graphene QDs and drug molecules, i.e., DOX and Curcumin along with their dual-application in targeted cancer therapy and tracking.[85] Copyright 2022 Elsevier Inc.

vaccine demonstrated potential anticancer efficacy and would enhance the survival time in the TUBO mice model.[79] A recent investigation conducted by Ahmed et al. has synthesized a Toll-like receptor 9 (TLR9) and CpG oligonucleotide (CpG ON) that were lyophilized and linked to streptavidin. These streptavidin-bound peptides were then conjugated with poly lactic-*co*-glycolic acid (PLGA) (Figure 12). Furthermore, the *in vitro* and *in vivo* studies were conducted on murine melanoma cell lines (B16.F10) and C57BL/6 and BALB/c mice, respectively. Their results demonstrated efficient application as a cancer vaccine in prostate cancer cell models.[81,93]

Yakati et al. demonstrated PLGA NP conjugation with tumor targeting peptides: CPKSNNGVC (CPK) and paclitaxel (PTX). Herein, the encapsulation took place using the emulsion solvent evaporation method, and the association between CPK peptide and PLGA NP was achieved via maleimide−thiol chemistry. The formed prodrug exhibited a high affinity for the monocarboxylate transporter 1 (MCT1) receptor, which mediated endocytosis selectively in CRC cells due to overexpression of MCT1 receptors. To further confirm the therapeutic potential of the nanopeptide drug, an assessment of angiogenesis was carried out in chick embryos. They observed a remarkable reduction in angiogenesis, i.e., new blood vessel formation. Therefore, it exhibits immense potential as antineoplastic nanodrugs specific to CRC.[84] A recent study conducted by Ghafary et al., adopted a hydrothermal and chemical synthesis route for the fabrication of MiRGD peptide (a chimeric peptide with Cys-Arg-Gly-Asp-

https://doi.org/10.1021/acsomega.2c03974
ACS Omega XXXX, XXX, XXX−XXX

Lys-Gly-Pro-Asp-Cys; CRGDKGPDC motif) and functionalized it with graphene QDs. These are then conjugated with cytostatic drugs such as DOX, cisplatin, curcumin, etc. as carrier moieties (Figure 13). However, for their study, they selected DOX and curcumin as hydrophobic and hydrophilic cytostatic drugs, respectively. On the basis of the qualitative and quantitative analysis, the formed peptide-f-NP drug exhibited enhanced and targeted drug delivery as well as tracking of tumor cells via *in vivo* studies conducted in mice.[85]

5. CHALLENGES AND OUTLOOK

The expanding literature demonstrates intricate nanoassemblies and biomolecule (nanobio) interactions, heretofore, the interceding for therapeutic and clinical applications are still in the exploratory stage.[94] For the development of peptide-f-NP drug conjugates with efficient potential for clinical manifestation, several important criteria need to be taken into consideration such as composition, target affinity, dynamic properties, stability, distribution, permeability, retention, cytotoxicity, and biocompatibility.[24] These criteria are arduous for evaluating nanobio interactions. Evaluation of the NPs for basic biological processes, including absorption, distribution, metabolism, excretion, and toxicity (ADMET) play crucial insights into the development of novel nanopeptide drugs.[94] Henceforth, effective strategies are needed for the synthesis of peptide molecules with therapeutic ability.

The peptides are sensitive to elimination from undesired *in vivo* interactions due to their short *in vivo* half-life and limited stability. A method to enhance its stability has been through the formation of self-assembled peptide-f-NPs. Ionic, H-bonds, and $\pi-\pi$ interactions are the major interactions in these nanostructures, with $\pi-\pi$ interactions playing a key role in the production of peptide-f-NPs. Unfortunately, progress is slow in this sector, resulting in a significant lack of understanding of these structures and their interconnections.[26]

Another important obstacle is the circulation time of peptides which is a major drawback for peptide synthesis. This is caused by the proteolytic cleavage of these therapeutic peptides. Therefore, studies based on enzymatic stability, profiling of peptide degradation, and pharmacokinetic and *in vivo* studies are crucial for further innovation and clinical intervention. To improve the circulation time of peptides for therapeutic intervention, certain approaches such as modification of peptides can be considered effective. For example, N- and C-terminal modifications like acetylation and amidation, respectively, could be adopted. Another aspect could be peptide cyclization which would reduce proteolytic degradation and thereby increase circulation time.[24]

Another facet is the accurate dosage of the formulated peptide-f-NP drugs. Due to the nanobio interactions in peptide-f-NP drugs, the nature of the drug molecule is heterogeneous. Thereby, the evaluation of precise dosage is crucial to elucidate various characteristics of a proposed peptide-f-NP drug. In addition to this, several parameters such as surface area and molar concentration also influence the accurate assessment of such drugs. Unfortunately, the obstacles for quantification are progressively increasing as the surface area has gone from single particle to complex nanoassemblies.[94−97]

Furthermore, the intervention of NPs for peptide-based drugs would also enhance the stability and biocompatibility and would be an excellent source for delivering mechanisms.[1] Henceforth, the amalgamation of NPs into peptide-derived therapeutics for anticancer activity would benefit clinical translation in the field of cancer treatment.

6. CONCLUSION

With advances in the understanding of cancer biology and clinical intervention, several strategies for peptide-based nanotherapeutics have tremendously increased and emerged as a potential role in oncology research. However, only a handful of nanopeptide drugs have made it to clinical trials. Our current research lacks a complete understanding of peptide-NP interactions and the various parameters (peptide length, amino acid sequence, NP size, surface chemistry, etc.) involved in the functionalization of peptide-NPs and therefore needs many detailed studies and vigorous analysis. This issue is thought to be due to the certainty that each NP-peptide interaction is exclusive, necessitating detailed individual evaluations for each of these superstructures. Furthermore, examination of only one of these interactions does not allow for generalization to other types of peptide-NP superstructures. Additionally, studies have to be extended toward an analysis of the relationship between the primary amino acid structures of the peptide and the binding NPs, which is an essentially difficult task that remains a huge challenge to current peptide therapeutics. Therefore, this continuously expanding universe of peptide-f-NPs needs more detailed studies and vigorous analysis. The lack of reliable and advanced spectroscopic and computational techniques also hinders our ability to fundamentally understand these complex nanostructures. Consequently, only a limited number of peptide-f-NP therapeutics have been FDA approved for cancer therapy. However, due to the urgent need for a medication to combat the steadily rising number of cancer mortality worldwide, the integration of nanostructured platforms for peptide-based drugs is poised to grab researchers' attention for the development of new anticancer therapeutics. Thus, the new innovative combinatorial intervention of NPs and peptide drugs would overcome the demerits of conventional cancer therapy strategies and, therefore, fast-track the clinical translation of peptide-f-NP nanotherapy for anticancer treatment.

AUTHOR INFORMATION

Corresponding Authors

Akanksha Gupta − *Department of Chemistry, Sri Venkateswara College, University of Delhi, Delhi 110007, India*; orcid.org/0000-0003-1398-2085; Email: akankshashachem05@gmail.com

Vijay Kumar Goel − *School of Physical Science, Jawaharlal Nehru University, Delhi 110067, India*; Email: vijaykgoel@jnu.ac.in

Vinod Kumar − *Special Centre for Nano Sciences, Jawaharlal Nehru University, Delhi 110067, India*; orcid.org/0000-0002-8362-8009; Email: kumarv@mail.jnu.ac.in

Authors

Ritika Sharma − *Department of Biochemistry, University of Delhi, Delhi 110021, India*

Shikha Jyoti Borah − *Special Centre for Nano Sciences, Jawaharlal Nehru University, Delhi 110067, India*

Bhawna − *Department of Chemistry, University of Delhi, Delhi 110007, India*

https://doi.org/10.1021/acsomega.2c03974
ACS Omega XXXX, XXX, XXX−XXX

Sanjeev Kumar − *Department of Chemistry and Department of Chemistry, Kirori Mal College, University of Delhi, Delhi 110007, India*
Poonam Singh − *Department of Applied Chemistry, Delhi Technological University, Delhi 110042, India*
Ravinder Kumar − *Department of Chemistry, Gurukula Kangri (Deemed to be University), Haridwar 249404 Uttarakhand, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c03974

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

S.J.B. and V.K.G. thank Jawaharlal Nehru University, New Delhi, and Sanganeria Foundation, New Delhi, for financial support, respectively. B. and S.K. thank University Grant Commission and Council of Scientific & Industrial Research (File no. 08/694(0004)/2018-EMR-I), respectively, for the Senior Research Fellowships.

■ REFERENCES

(1) Guo, R.-C.; Zhang, X.-H.; Ji, L.; Wei, Z.-J.; Duan, Z.-Y.; Qiao, Z.-Y.; Wang, H. Recent progress of therapeutic peptide based nanomaterials: from synthesis and self-assembly to cancer treatment. *Biomater. Sci.* **2020**, *8* (22), 6175−6189.
(2) Moasses Ghafary, S.; Rahimjazi, E.; Hamzehil, H.; Modarres Mousavi, S. M.; Nikkhah, M.; Hosseinkhani, S. Design and preparation of a Theranostic Peptideticle for targeted Cancer therapy: Peptide-based Codelivery of doxorubicin/curcumin and graphene quantum dots. *Nanomed. Nanotechnol. Biol. Med.* **2022**, *42*, 102544.
(3) Mehrotra, N.; Kharbanda, S.; Singh, H. Peptide-based combination nanoformulations for cancer therapy. *Nanomed.* **2020**, *15* (22), 2201−2217.
(4) Delfi, M.; Sartorius, R.; Ashrafizadeh, M.; Sharifi, E.; Zhang, Y.; De Berardinis, P.; Zarrabi, A.; Varma, R. S.; Tay, F. R.; Smith, B. R.; Makvandi, P. Self-assembled peptide and protein nanostructures for anti-cancer therapy: Targeted delivery, stimuli-responsive devices and immunotherapy. *Nano Today* **2021**, *38*, 101119.
(5) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **1991**, *354* (6348), 82−84.
(6) Klimpel, A.; Lützenburg, T.; Neundorf, I. Recent advances of anti-cancer therapies including the use of cell-penetrating peptides. *Curr. Opin. Pharmacol.* **2019**, *47*, 8−13.
(7) Fjell, C. D.; Hiss, J. A.; Hancock, R. E.; Schneider, G. Designing antimicrobial peptides: form follows function. *Nat. Rev. Drug Discovery* **2012**, *11* (1), 37−51.
(8) Xie, M.; Liu, D.; Yang, Y. Anti-cancer peptides: Classification, mechanism of action, reconstruction and modification. *Open Biol.* **2020**, *10* (7), 200004.
(9) Trac, N. T.; Chung, E. J. Peptide-based targeting of immunosuppressive cells in cancer. *Bioact. Mater.* **2020**, *5* (1), 92−101.
(10) Zou, R.; Wang, Q.; Wu, J.; Wu, J.; Schmuck, C.; Tian, H. Peptide self-assembly triggered by metal ions. *Chem. Soc. Rev.* **2015**, *44* (15), 5200−5219.
(11) Dong, R.; Zhou, Y.; Huang, X.; Zhu, X.; Lu, Y.; Shen, J. Functional supramolecular polymers for biomedical applications. *Adv. Mater.* **2015**, *27* (3), 498−526.
(12) Ulijn, R. V.; Smith, A. M. Designing peptide based nanomaterials. *Chem. Soc. Rev.* **2008**, *37* (4), 664−675.
(13) Sharma, R.; Bhawna; Kumar, S.; Singh, P.; Gupta, A.; Kumar, V. Layered Double Hydroxide Nanomaterials: Biomedical Applications, Current Status and Challenges. *Nano Life* **2021**, *11* (3), 2130008.
(14) Sharma, R.; Verma, B.; Kumar, S.; Gupta, A.; Sahu, P. K.; Singh, P.; Kumar, V. Recent updates on applications of ionic liquids (ILs) for biomedical sciences. *Journal of the Iranian Chemical Society* **2022**, *19*, 3215−3228.
(15) Kumar, S.; Sharma, R.; Bhawna; Gupta, A.; Singh, P.; Kalia, S.; Thakur, P.; Kumar, V. Prospects of Biosensors Based on Functionalized and Nanostructured Solitary Materials: Detection of Viral Infections and Other Risks. *ACS Omega* **2022**, *7* (26), 22073−22088.
(16) Sharma, R.; Gupta, A.; Kumar, R.; Singh, P.; Kumar, V. An Update on COVID-19: Role of Nanotechnology in Vaccine Development. *SMC Bulletin* **2020**, *11* (2), 88−96.
(17) Coppage, R.; Slocik, J. M.; Ramezani-Dakhel, H.; Bedford, N. M.; Heinz, H.; Naik, R. R.; Knecht, M. R. Exploiting localized surface binding effects to enhance the catalytic reactivity of peptide-capped nanoparticles. *J. Am. Chem. Soc.* **2013**, *135* (30), 11048−11054.
(18) Sarikaya, M.; Somerman, M.; Tammerler-Behar, C.; Hanson, F.; Zhang, H.; Gungormus, M. *Reagents and methods for treating dental disease*, U.S. Patent US9809633B2, 2012
(19) Tarvirdipour, S.; Huang, X.; Mihali, V.; Schoenenberger, C.-A.; Palivan, C. G. Peptide-based nanoassemblies in gene therapy and diagnosis: paving the way for clinical application. *Molecules* **2020**, *25* (15), 3482.
(20) Zhang, L.; Huang, Y.; Lindstrom, A. R.; Lin, T.-Y.; Lam, K. S.; Li, Y. Peptide-based materials for cancer immunotherapy. *Theranostics* **2019**, *9* (25), 7807−7825.
(21) Singh, A.; Deshpande, N.; Pramanik, N.; Jhunjhunwala, S.; Rangarajan, A.; Atreya, H. S. Optimized peptide based inhibitors targeting the dihydrofolate reductase pathway in cancer. *Sci. Rep.* **2018**, *8* (1), 3190.
(22) Jeong, W.-J.; Bu, J.; Kubiatowicz, L. J.; Chen, S. S.; Kim, Y.; Hong, S. Peptide−nanoparticle conjugates: a next generation of diagnostic and therapeutic platforms? *Nano Converg.* **2018**, *5* (1), 1−18.
(23) Rong, L.; Lei, Q.; Zhang, X. Z. Recent advances on peptide-based theranostic nanomaterials. *View* **2020**, *1* (4), 20200050.
(24) Ayo, A.; Laakkonen, P. Peptide-based strategies for targeted tumor treatment and imaging. *Pharmaceutics* **2021**, *13* (4), 481.
(25) Panda, J. J.; Chauhan, V. S. Short peptide based self-assembled nanostructures: implications in drug delivery and tissue engineering. *Polym. Chem.* **2014**, *5* (15), 4418−4436.
(26) Habibi, N.; Kamaly, N.; Memic, A.; Shafiee, H. Self-assembled peptide-based nanostructures: Smart nanomaterials toward targeted drug delivery. *Nano today* **2016**, *11* (1), 41−60.
(27) Shi, Y.; Lu, A.; Wang, X.; Belhadj, Z.; Wang, J.; Zhang, Q. A review of existing strategies for designing long-acting parenteral formulations: focus on underlying mechanisms, and future perspectives. *Acta Pharm. Sin. B* **2021**, *11* (8), 2396−2415.
(28) Li, X.; Liu, L.; Fu, Y.; Chen, H.; Abualrejal, M. M.; Zhang, H.; Wang, Z.; Zhang, H. Peptide-enhanced tumor accumulation of upconversion nanoparticles for sensitive upconversion luminescence/magnetic resonance dual-mode bioimaging of colorectal tumors. *Acta Biomater.* **2020**, *104*, 167−175.
(29) Liu, L.; Li, X.; Zhang, H.; Chen, H.; Abualrejal, M. M.; Song, D.; Wang, Z. Six-in-one peptide functionalized upconversion@polydopamine nanoparticle-based ratiometric fluorescence sensing platform for real-time evaluating anticancer efficacy through monitoring caspase-3 activity. *Sens. Actuators B Chem.* **2021**, *333*, 129554.
(30) Yu, C.-Y.; Yuan, Z.; Cao, Z.; Wang, B.; Qiao, C.; Li, J.; Xiao, X. A muscle-targeting peptide displayed on AAV2 improves muscle tropism on systemic delivery. *Gene Ther.* **2009**, *16* (8), 953−962.
(31) Diaconu, I.; Denby, L.; Pesonen, S.; Cerullo, V.; Bauerschmitz, G. J.; Guse, K.; Rajecki, M.; Dias, J. D.; Taari, K.; Kanerva, A.; Baker, A. H.; Hemminki, A. Serotype chimeric and fiber-mutated adenovirus Ad5/19p-HIT for targeting renal cancer and untargeting the liver. *Hum. Gene Ther.* **2009**, *20* (6), 611−620.

(32) Pasqualini, R.; Ruoslahti, E. Organ targeting in vivo using phage display peptide libraries. *Nature* **1996**, *380* (6572), 364−366.
(33) Altman, M.; Lee, P.; Rich, A.; Zhang, S. Conformational behavior of ionic self-complementary peptides. *Protein Sci.* **2000**, *9* (6), 1095−1105.
(34) Mullard, A. 2020 FDA drug approvals. *Nat. Rev. Drug Discovery* **2021**, *20* (2), 85−91.
(35) Ayoub, M.; Scheidegger, D. Peptide drugs, overcoming the challenges, a growing business. *Chimica oggi* **2006**, *24* (4), 46.
(36) Merrifield, R. B. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149−2154.
(37) Li, Y.; Tang, Z.; Prasad, P. N.; Knecht, M. R.; Swihart, M. T. Peptide-mediated synthesis of gold nanoparticles: effects of peptide sequence and nature of binding on physicochemical properties. *Nanoscale* **2014**, *6* (6), 3165−3172.
(38) Colangelo, E.; Chen, Q.; Davidson, A. M.; Paramelle, D.; Sullivan, M. B.; Volk, M.; LéVy, R. L. Computational and experimental investigation of the structure of peptide monolayers on gold nanoparticles. *Langmuir* **2017**, *33* (1), 438−449.
(39) Hu, B.; Kong, F.; Gao, X.; Jiang, L.; Li, X.; Gao, W.; Xu, K.; Tang, B. Avoiding Thiol Compound Interference: A Nanoplatform Based on High-Fidelity Au−Se Bonds for Biological Applications. *Angew. Chem.* **2018**, *130* (19), 5404−5407.
(40) Luan, M.; Shi, M.; Pan, W.; Li, N.; Tang, B. A gold−selenium-bonded nanoprobe for real-time in situ imaging of the upstream and downstream relationship between uPA and MMP-9 in cancer cells. *Chem. Commun.* **2019**, *55* (41), 5817−5820.
(41) Chen, H.; Li, X.; Liu, F.; Zhang, H.; Wang, Z. Renal clearable peptide functionalized NaGdF4 nanodots for high-efficiency tracking orthotopic colorectal tumor in mouse. *Mol. Pharmaceutics* **2017**, *14* (9), 3134−3141.
(42) Liu, F.; He, X.; Zhang, J.; Zhang, H.; Wang, Z. Employing tryptone as a general phase transfer agent to produce renal clearable nanodots for bioimaging. *Small* **2015**, *11* (30), 3676−3685.
(43) Upert, G.; Bouillère, F.; Wennemers, H. Oligoprolines as scaffolds for the formation of silver nanoparticles in defined sizes: correlating molecular and nanoscopic dimensions. *Angew. Chem., Int. Ed.* **2012**, *51* (17), 4231−4234.
(44) Graf, P.; Mantion, A.; Foelske, A.; Shkilnyy, A.; Mašić, A.; Thünemann, A. F.; Taubert, A. Peptide-coated silver nanoparticles: synthesis, surface chemistry, and pH-triggered, reversible assembly into particle assemblies. *Chem.—Eur. J.* **2009**, *15* (23), 5831−5844.
(45) Si, S.; Mandal, T. K. Tryptophan-based peptides to synthesize gold and silver nanoparticles: a mechanistic and kinetic study. *Chem.—Eur. J.* **2007**, *13* (11), 3160−3168.
(46) Roma-Rodrigues, C.; Heuer-Jungemann, A.; Fernandes, A. R.; Kanaras, A. G.; Baptista, P. V. Peptide-coated gold nanoparticles for modulation of angiogenesis in vivo. *Int. J. Nanomedicine* **2016**, *11*, 2633.
(47) Wilder, L. M.; Fies, W. A.; Rabin, C.; Webb, L. J.; Crooks, R. M. Conjugation of an α-Helical Peptide to the Surface of Gold Nanoparticles. *Langmuir* **2019**, *35* (9), 3363−3371.
(48) Bartczak, D.; Kanaras, A. G. Preparation of peptide-functionalized gold nanoparticles using one pot EDC/sulfo-NHS coupling. *Langmuir* **2011**, *27* (16), 10119−10123.
(49) Fu, Y.; Li, X.; Chen, H.; Wang, Z.; Yang, W.; Zhang, H. CXC chemokine receptor 4 antagonist functionalized renal clearable manganese-doped iron oxide nanoparticles for active-tumor-targeting magnetic resonance imaging-guided bio-photothermal therapy. *ACS Appl. Bio Mater.* **2019**, *2* (8), 3613−3621.
(50) Li, X.; Jian, M.; Sun, Y.; Zhu, Q.; Wang, Z. The Peptide Functionalized Inorganic Nanoparticles for Cancer-Related Bioanalytical and Biomedical Applications. *Molecules* **2021**, *26* (11), 3228.
(51) De Castro, R. R.; Do Carmo, F. A.; Martins, C.; Simon, A.; De Sousa, V. P.; Rodrigues, C. R.; Cabral, L. M.; Sarmento, B. Clofazimine functionalized polymeric nanoparticles for brain delivery in the tuberculosis treatment. *Int. J. Pharm.* **2021**, *602*, 120655.
(52) Joshi, S.; Sharma, P.; Siddiqui, R.; Kaushal, K.; Sharma, S.; Verma, G.; Saini, A. A review on peptide functionalized graphene derivatives as nanotools for biosensing. *Microchimica Acta* **2020**, *187* (1), 1−15.
(53) Feng, T.; Feng, D.; Shi, W.; Li, X.; Ma, H. A graphene oxide-peptide fluorescence sensor for proteolytically active prostate-specific antigen. *Mol. Biosyst.* **2012**, *8* (5), 1441−1445.
(54) Ding, X.; Wang, Y.; Cheng, W.; Mo, F.; Sang, Y.; Xu, L.; Ding, S. Aptamer based electrochemical adenosine triphosphate assay based on a target-induced dendritic DNA nanoassembly. *Microchimica Acta* **2017**, *184* (2), 431−438.
(55) Yang, H.; Fung, S. Y.; Sun, W.; Mikkelsen, S.; Pritzker, M.; Chen, P. Ionic-complementary peptide-modified highly ordered pyrolytic graphite electrode for biosensor application. *Biotechnology progress* **2008**, *24* (4), 964−971.
(56) Levine, R. M.; Scott, C. M.; Kokkoli, E. Peptide functionalized nanoparticles for nonviral gene delivery. *Soft Matter* **2013**, *9* (4), 985−1004.
(57) Li, X.; Li, Y.; Qiu, Q.; Wen, Q.; Zhang, Q.; Yang, W.; Yuwen, L.; Weng, L.; Wang, L. Efficient biofunctionalization of MoS2 nanosheets with peptides as intracellular fluorescent biosensor for sensitive detection of caspase-3 activity. *J. Colloid Interface Sci.* **2019**, *543*, 96−105.
(58) Zhu, L.; Zhao, H.; Zhou, Z.; Xia, Y.; Wang, Z.; Ran, H.; Li, P.; Ren, J. Peptide-functionalized phase-transformation nanoparticles for low intensity focused ultrasound-assisted tumor imaging and therapy. *Nano Lett.* **2018**, *18* (3), 1831−1841.
(59) Lin, B.; Wu, J.; Wang, Y.; Sun, S.; Yuan, Y.; Tao, X.; Lv, R. Peptide functionalized upconversion/NIR II luminescent nanoparticles for targeted imaging and therapy of oral squamous cell carcinoma. *Biomater. Sci.* **2021**, *9* (3), 1000−1007.
(60) De La Fuente, J. M.; Berry, C. C. Tat peptide as an efficient molecule to translocate gold nanoparticles into the cell nucleus. *Bioconjugate Chem.* **2005**, *16* (5), 1176−1180.
(61) Li, R.; Gao, R.; Wang, Y.; Liu, Z.; Xu, H.; Duan, A.; Zhang, F.; Ma, L. Gastrin releasing peptide receptor targeted nano-graphene oxide for near-infrared fluorescence imaging of oral squamous cell carcinoma. *Sci. Rep.* **2020**, *10* (1), 11434.
(62) Campos, E. V.; Pereira, A. E.; De Oliveira, J. L.; Carvalho, L. B.; Guilger-Casagrande, M.; De Lima, R.; Fraceto, L. F. How can nanotechnology help to combat COVID-19? Opportunities and urgent need. *J. Nanobiotechnol.* **2020**, *18* (1), 1−23.
(63) Khalil, I. A.; Harashima, H. An efficient PEGylated gene delivery system with improved targeting: Synergism between octaarginine and a fusogenic peptide. *Int. J. Pharm.* **2018**, *538* (1−2), 179−187.
(64) Shtykalova, S.; Egorova, A.; Maretina, M.; Baranov, V.; Kiselev, A. Magnetic Nanoparticles as a Component of Peptide-Based DNA Delivery System for Suicide Gene Therapy of Uterine Leiomyoma. *Bioengineering* **2022**, *9* (3), 112.
(65) Araste, F.; Abnous, K.; Hashemi, M.; Taghdisi, S. M.; Ramezani, M.; Alibolandi, M. Peptide-based targeted therapeutics: Focus on cancer treatment. *J. Controlled Release* **2018**, *292*, 141−162.
(66) Liu, X.; Wu, F.; Ji, Y.; Yin, L. Recent advances in anti-cancer protein/peptide delivery. *Bioconjugate Chem.* **2019**, *30* (2), 305−324.
(67) Li, Y.; Zhang, M.; Han, H.; Zhang, B.; Matson, J. B.; Chen, D.; Li, W.; Wang, Y. Peptide-based supramolecular photodynamic therapy systems: From rational molecular design to effective cancer treatment. *Chem. Eng. J.* **2022**, *436*, 135240.
(68) Liu, W.; Tang, H.; Li, L.; Wang, X.; Yu, Z.; Li, J. Peptide-based therapeutic cancer vaccine: Current trends in clinical application. *Cell Prolif.* **2021**, *54* (5), No. e13025.
(69) Xiang, S. D.; Wilson, K. L.; Goubier, A.; Heyerick, A.; Plebanski, M. Design of peptide-based nanovaccines targeting leading antigens from gynecological cancers to induce HLA-A2. 1 restricted CD8+ T cell responses. *Front. Immunol.* **2018**, *9*, 1.
(70) Kanazawa, T.; Morisaki, K.; Suzuki, S.; Takashima, Y. Prolongation of life in rats with malignant glioma by intranasal siRNA/drug codelivery to the brain with cell-penetrating peptide-modified micelles. *Mol. Pharmaceutics* **2014**, *11* (5), 1471−1478.

https://doi.org/10.1021/acsomega.2c03974
ACS Omega XXXX, XXX, XXX−XXX

(71) Lu, L.; Zhang, H.; Zhou, Y.; Lin, J.; Gao, W.; Yang, T.; Jin, J.; Zhang, L.; Nagle, D. G.; Zhang, W.; Wu, Y.; Chen, H.; Luan, X. Polymer chimera of stapled oncolytic peptide coupled with anti-PD-L1 peptide boosts immunotherapy of colorectal cancer. *Theranostics* **2022**, *12* (7), 3456−3473.
(72) Kumar, A.; Ma, H.; Zhang, X.; Huang, K.; Jin, S.; Liu, J.; Wei, T.; Cao, W.; Zou, G.; Liang, X.-J. Gold nanoparticles functionalized with therapeutic and targeted peptides for cancer treatment. *Biomaterials* **2012**, *33* (4), 1180−1189.
(73) Akrami, M.; Balalaie, S.; Hosseinkhani, S.; Alipour, M.; Salehi, F.; Bahador, A.; Haririan, I. Tuning the anticancer activity of a novel pro-apoptotic peptide using gold nanoparticle platforms. *Sci. Rep.* **2016**, *6* (1), 1−12.
(74) Shim, M. K.; Moon, Y.; Yang, S.; Kim, J.; Cho, H.; Lim, S.; Yoon, H. Y.; Seong, J.-K.; Kim, K. Cancer-specific drug-drug nanoparticles of pro-apoptotic and cathepsin B-cleavable peptide-conjugated doxorubicin for drug-resistant cancer therapy. *Biomaterials* **2020**, *261*, 120347.
(75) Rousselle, C.; Clair, P.; Lefauconnier, J.-M.; Kaczorek, M.; Scherrmann, J.-M.; Temsamani, J. New advances in the transport of doxorubicin through the blood-brain barrier by a peptide vector-mediated strategy. *Mol. Pharmacol.* **2000**, *57* (4), 679−686.
(76) Moon, Y.; Shim, M. K.; Choi, J.; Yang, S.; Kim, J.; Yun, W. S.; Cho, H.; Park, J. Y.; Kim, Y.; Seong, J.-K.; Kim, K. Anti-PD-L1 peptide-conjugated prodrug nanoparticles for targeted cancer immunotherapy combining PD-L1 blockade with immunogenic cell death. *Theranostics* **2022**, *12* (5), 1999.
(77) Hazeri, Y.; Samie, A.; Ramezani, M.; Alibolandi, M.; Yaghoobi, E.; Dehghani, S.; Zolfaghari, R.; Khatami, F.; Zavvar, T.; Nameghi, M. A.; Abnous, K.; Taghdisi, S. M. Dual-targeted delivery of doxorubicin by mesoporous silica nanoparticle coated with AS1411 aptamer and RGDK-R peptide to breast cancer in vitro and in vivo. *J. Drug Delivery Sci. Technol.* **2022**, *71*, 103285.
(78) Li, J.; Wu, S.; Tian, X.; Li, X. Fabrication of a multifunctional nanomaterial from a mussel-derived peptide for multimodal synergistic cancer therapy. *Chem. Eng. J.* **2022**, *446*, 136837.
(79) Farzad, N.; Barati, N.; Momtazi-Borojeni, A. A.; Yazdani, M.; Arab, A.; Razazan, A.; Shariat, S.; Mansourian, M.; Abbasi, A.; Saberi, Z.; Badiee, A.; Jalali, S. A.; Jaafari, M. R. P435 HER2/neu-derived peptide conjugated to liposomes containing DOPE as an effective prophylactic vaccine formulation for breast cancer. *Artif. Cells Nanomed. Biotechnol.* **2019**, *47* (1), 664−672.
(80) Rezvantalab, S.; Drude, N. I.; Moraveji, M. K.; Güvener, N.; Koons, E. K.; Shi, Y.; Lammers, T.; Kiessling, F. PLGA-based nanoparticles in cancer treatment. *Front. Pharmacol.* **2018**, *9*, 1260.
(81) Ahmed, K. K.; Geary, S. M.; Salem, A. K. Surface engineering tumor cells with adjuvant-loaded particles for use as cancer vaccines. *J. Controlled Release* **2017**, *248*, 1−9.
(82) Wang, H.; Yan, Y.-Q.; Yi, Y.; Wei, Z.-Y.; Chen, H.; Xu, J.-F.; Wang, H.; Zhao, Y.; Zhang, X. Supramolecular peptide therapeutics: Host−guest interaction-assisted systemic delivery of anticancer peptides. *CCS Chemistry* **2020**, *2* (6), 739−748.
(83) Ren, C.; Wang, Z.; Zhang, X.; Gao, J.; Gao, Y.; Zhang, Y.; Liu, J.; Yang, C.; Liu, J. Construction of all-in-one peptide nanomedicine with photoacoustic imaging guided mild hyperthermia for enhanced cancer chemotherapy. *Chem. Eng. J.* **2021**, *405*, 127008.
(84) Yakati, V.; Vangala, S.; Madamsetty, V. S.; Banerjee, R.; Moku, G. Enhancing the anticancer effect of paclitaxel by using polymeric nanoparticles decorated with colorectal cancer targeting CPKSNNGVC-peptide. *J. Drug Delivery Sci. Technol.* **2022**, *68*, 103125.
(85) Moasses Ghafary, S.; Rahimjazi, E.; Hamzehil, H.; Modarres Mousavi, S. M.; Nikkhah, M.; Hosseinkhani, S. Design and preparation of a Theranostic Peptideticle for targeted Cancer therapy: Peptide-based Codelivery of doxorubicin/curcumin and graphene quantum dots. *Nanomed. Nanotechnol. Biol. Med.* **2022**, *42*, 102544.
(86) Xiang, S. D.; Wilson, K. L.; Goubier, A.; Heyerick, A.; Plebanski, M. Design of peptide-based nanovaccines targeting leading antigens from gynecological cancers to induce HLA-A2. 1 restricted CD8+ T cell responses. *Front. Immunol.* **2018**, *9*, 2968.
(87) Stephenson, J. M.; Banerjee, S.; Saxena, N. K.; Cherian, R.; Banerjee, S. K. Neuropilin-1 is differentially expressed in myoepithelial cells and vascular smooth muscle cells in preneoplastic and neoplastic human breast: a possible marker for the progression of breast cancer. *Int. J. Cancer* **2002**, *101* (5), 409−414.
(88) Momand, J.; Wu, H.-H.; Dasgupta, G. MDM2—master regulator of the p53 tumor suppressor protein. *Gene* **2000**, *242* (1−2), 15−29.
(89) Vousden, K. H.; Lu, X. Live or let die: the cell's response to p53. *Nat. Rev. Cancer* **2002**, *2* (8), 594−604.
(90) Sugahara, K. N.; Teesalu, T.; Karmali, P. P.; Kotamraju, V. R.; Agemy, L.; Girard, O. M.; Hanahan, D.; Mattrey, R. F.; Ruoslahti, E. Tissue-penetrating delivery of compounds and nanoparticles into tumors. *Cancer Cell* **2009**, *16* (6), 510−520.
(91) Bielenberg, D. R.; Pettaway, C. A.; Takashima, S.; Klagsbrun, M. Neuropilins in neoplasms: expression, regulation, and function. *Exp. Cell Res.* **2006**, *312* (5), 584−593.
(92) Pazgier, M.; Liu, M.; Zou, G.; Yuan, W.; Li, C.; Li, C.; Li, J.; Monbo, J.; Zella, D.; Tarasov, S. G.; Lu, W. Structural basis for high-affinity peptide inhibition of p53 interactions with MDM2 and MDMX. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (12), 4665−4670.
(93) Rezvantalab, S.; Drude, N. I.; Moraveji, M. K.; Güvener, N.; Koons, E. K.; Shi, Y.; Lammers, T.; Kiessling, F. PLGA-based nanoparticles in cancer treatment. *Front. Pharmacol.* **2018**, *9*, 1260.
(94) Liu, J.; Guo, M.; Chen, C. Nano-bio interactions: a major principle in the dynamic biological processes of nano-assemblies. *Adv. Drug Delivery Rev.* **2022**, *186*, 114318.
(95) Malik, S.; Kumar, V.; Liu, C. H.; Shih, K. C.; Krueger, S.; Nieh, M. P.; Bahal, R. Head on Comparison of Self-and Nano-Assemblies of Gamma Peptide Nucleic Acid Amphiphiles. *Adv. Funct. Mater.* **2022**, *32* (7), 2109552.
(96) Liu, Y.; Sun, Y.; Zhang, W. Synthesis of Stimuli-Responsive Block Copolymers and Block Copolymer Nano-assemblies. *Chin. J. Chem.* **2022**, *40* (8), 965−972.
(97) Sarkar, A. K.; Debnath, K.; Arora, H.; Seth, P.; Jana, N. R.; Jana, N. R. Direct Cellular Delivery of Exogenous Genetic Material and Protein via Colloidal Nano-Assemblies with Biopolymer. *ACS Appl. Mater. Interfaces* **2022**, *14* (2), 3199−3206.

https://doi.org/10.1021/acsomega.2c03974
ACS Omega XXXX, XXX, XXX−XXX